

Digital Personalized Health and Medicine

Proceedings of MIE 2020



Editors: Louise B. Pape-Haugaard
Christian Lovis
Inge Cort Madsen
Patrick Weber
Per Hostrup Nielsen
Philip Scott

Digital health and medical informatics have grown in importance in recent years, and have now become central to the provision of effective healthcare around the world.

This book presents the proceedings of the 30th Medical Informatics Europe conference (MIE). This edition of the conference, hosted by the European Federation for Medical Informatics (EFMI) since the 1970s, was due to be held in Geneva, Switzerland in April 2020, but as a result of measures to prevent the spread of the Covid19 pandemic, the conference itself had to be cancelled.

Nevertheless, because this collection of papers offers a wealth of knowledge and experience across the full spectrum of digital health and medicine, it was decided to publish the submissions accepted in the review process and confirmed by the Scientific Program Committee for publication, and these are published here as planned. The 232 papers are themed under 6 section headings: biomedical data, tools and methods; supporting care delivery; health and prevention; precision medicine and public health; human factors and citizen centered digital health; and ethics, legal and societal aspects. A 7th section deals with the Swiss personalized health network, and section 8 includes the 125 posters accepted for the conference.

Offering an overview of current trends and developments in digital health and medical informatics, the book provides a valuable information resource for researchers and health practitioners alike.



ISBN 978-1-64368-082-8 (print)

ISBN 978-1-64368-083-5 (online)

ISSN 0926-9630 (print)

ISSN 1879-8365 (online)

DIGITAL PERSONALIZED HEALTH AND MEDICINE

Studies in Health Technology and Informatics

International health informatics is driven by developments in biomedical technologies and medical informatics research that are advancing in parallel and form one integrated world of information and communication media and result in massive amounts of health data. These components include genomics and precision medicine, machine learning, translational informatics, intelligent systems for clinicians and patients, mobile health applications, data-driven telecommunication and rehabilitative technology, sensors, intelligent home technology, EHR and patient-controlled data, and Internet of Things.

Studies in Health Technology and Informatics (HTI) series was started in 1990 in collaboration with EU programmes that preceded the Horizon 2020 to promote biomedical and health informatics research. It has developed into a highly visible global platform for the dissemination of original research in this field, containing more than 250 volumes of high-quality works from all over the world.

The international Editorial Board selects publications with relevance and quality for the field. All contributions to the volumes in the series are peer reviewed.

Volumes in the HTI series are submitted for indexing by MEDLINE/PubMed; Web of Science: Conference Proceedings Citation Index – Science (CPCI-S) and Book Citation Index – Science (BKCI-S); Google Scholar; Scopus; EMCare.

Series Editors:

B. Blobel, O. Bodenreider, E. Borycki, M. Braunstein, C. Bühler, J.P. Christensen, R. Cooper, R. Cornet, J. Dewen, O. Le Dour, P.C. Dykes, A. Famili, M. González-Sancho, E.J.S. Hovenga, J.W. Jutai, Z. Kolitsi, C.U. Lehmann, J. Mantas, V. Maojo, A. Moen, J.F.M. Molenbroek, G. de Moor, M.A. Musen, P.F. Niederer, C. Nøhr, A. Pedotti, N. Peek, O. Rienhoff, G. Riva, W. Rouse, K. Saranto, M.J. Scherer, S. Schürer, E.R. Siegel, C. Safran, N. Sarkar, T. Solomonides, E. Tam, J. Tenenbaum, B. Wiederhold, P. Wilson and L.H.W. van der Woude

Volume 270

Recently published in this series

- Vol. 269 R.A. Logan and E.R. Siegel (Eds.), Health Literacy in Clinical Practice and Public Health – New Initiatives and Lessons Learned at the Intersection with other Disciplines
- Vol. 268 A.J. Maeder, S. Champion, C. Moores and R. Golley (Eds.), Information Technology Based Methods for Health Behaviours – Selected Papers from Global Telehealth 2019
- Vol. 267 R. Röhrig, H. Binder, H.-U. Prokosch, U. Sax, I. Schmidtman, S. Stolpe and A. Zapf (Eds.), German Medical Data Sciences: Shaping Change – Creative Solutions for Innovative Medicine – Proceedings of the 64th Annual Meeting of the German Association of Medical Informatics, Biometry and Epidemiology (gmds e.V.) 2019 in Dortmund, Germany – GMDS 2019

ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

Digital Personalized Health and Medicine

Proceedings of MIE 2020

Edited by

Louise B. Pape-Haugaard

*Associate Professor, Department of Health Science and Technology, Medical
Informatics Group, Aalborg University, Denmark*

Christian Lovis

*Chair, Division of Medical Information Sciences, University Hospitals of
Geneva (HUG), Switzerland*

Inge Cort Madsen

Faculty of Health, VIA University College, Aarhus, Denmark

Patrick Weber

Nice Computing SA, Le Mont/Lausanne, Switzerland

Per Hostrup Nielsen

*Department of Cardiothoracic and Vascular Surgery, Aarhus University
Hospital, Aarhus, Denmark*

and

Philip Scott

Reader in Health Informatics, University of Portsmouth, UK

IOS
Press

Amsterdam • Berlin • Washington, DC

© 2020 European Federation for Medical Informatics (EFMI) and IOS Press.

This book is published online with Open Access and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

ISBN 978-1-64368-082-8 (print)

ISBN 978-1-64368-083-5 (online)

Library of Congress Control Number: 2020939284

doi: 10.3233/SHTI270

Publisher

IOS Press BV

Nieuwe Hemweg 6B

1013 BG Amsterdam

Netherlands

fax: +31 20 687 0019

e-mail: order@iospress.nl

For book sales in the USA and Canada:

IOS Press, Inc.

6751 Tepper Drive

Clifton, VA 20124

USA

Tel.: +1 703 830 6300

Fax: +1 703 830 2300

sales@iospress.com

LEGAL NOTICE

The publisher is not responsible for the use which might be made of the following information.

PRINTED IN THE NETHERLANDS

Digital Personalized Health and Medicine: Preface

Philip SCOTT^a and Louise B. PAPE-HAUGAARD^b

^a*School of Computing, University of Portsmouth, United Kingdom*

^b*Dept of Health Science and Technology, Aalborg University, Denmark*

This volume presents the proceedings of the 30th Medical Informatics Europe conference (MIE), organized in Geneva, Switzerland, in April 2020. This collection of papers offers a wealth of knowledge and experience across the full spectrum of digital health and medicine.

MIE conferences have been hosted by the European Federation for Medical Informatics (EFMI) since the 1970s. Over those decades, we have been privileged to share in a growing community of expertise in health and care informatics and seen the advancement of the field from a few pioneers to a diverse international network.

The overarching concept of Digital Personalized Health and Medicine is elaborated under the six programme themes as the structure of this volume:

- Biomedical data, tools and methods
- Supporting care delivery
- Health and prevention
- Precision medicine and public health
- Human factors and citizen centered digital health
- Ethics, legal and societal aspects

The topics of the conference proceedings demonstrate that crucial scientific work is still in progress across this scientific continuum and that much existing knowledge is yet to be widely adopted in routine practice. For example, recent years have seen unprecedented leaps forward in data science and machine learning, yet important work still continues on basic prerequisites such as data quality and computable semantic interoperability. Further excellent work continues in human-computer interaction, policy, workforce development, ethics and regulation. Increasing attention is being paid to citizen and patient concerns, whether it be trust in computerised clinical guidance, privacy and consent or issues of safety and accountability.

Helpfully, the Learning Health System is increasingly seen as a unifying concept for basic and translational health and care informatics, incorporating standards for representing biomedical knowledge, machine learning, precision medicine, clinical decision support systems, quality improvement and behaviour change.

We commend this body of papers to readers as a powerful educational resource.

MIE 2020 was organized in partnership with the World Health Organization (WHO), the International Telecommunication Union (ITU), the State of Geneva (Geneva), the University of Geneva (UNIGE) and the University Hospitals of Geneva (HUG). The local team would particularly like to acknowledge the work of Dr Vasiliki Foufi PhD and Christophe Gaudet-Blavignac BSc CS MMed. The editors would like to express thanks to our doctoral student assistants, Taiwo Adedeji MSc and Obinwa Ozonze MSc.

Louise B. Pape-Haugaard, Philip Scott
Editors

Christian Lovis, Inge Cort Madsen, Patrick Weber, Per Hostrup Nielsen
Scientific Programme Committee

Portsmouth/Aalborg, 10 March 2020

Digital Personalized Health and Medicine: COVID-19

Louise B. PAPE-HAUGAARD^a, Lăcrămioara STOICU-TIVADAR^b, Inge
CORT MADSEN^c, Alfred WINTER^d and Christian LOVIS^e

^a *Dept of Health Science and Technology, Aalborg University, Denmark*

^b *Department of Automation and Applied Informatics, University Politehnica
Timișoara, Timișoara, Romania*

^c *Faculty of Health, VIA University College, Aarhus, Denmark*

^d *Institute for Medical Informatics, Statistics and
Epidemiology (IMISE), University Leipzig, Leipzig, Germany*

^e *Chair, Division of Medical Information Sciences, University Hospitals of Geneva
(HUG), Switzerland*

***On behalf of EFMI, the Scientific Program Committee and the Editors of the
Proceedings of MIE2020***

1. Message from EFMI

In regard with the fast evolving situation due to the pandemic of coronavirus, the absolute necessity to slow down the progression of the spread of the virus to avoid collapse of the care systems, the State of Emergency has been decided by the Swiss Federal Council, implementing a locked down of the country. The global situation, has thus led to cancel the MIE2020 Conference.

MIE2020's primary concern is the safety of both participants, sponsors, committees and all of those who had planned to attend the Congress along with the investment that all have made to be present.

It has been decided to publish the Proceedings. Submissions accepted in the review process and confirmed by the SPC for publication will be published and indexed as planned.

The Editors of the Proceedings thank IOS Press for their support in this difficult situation.

This page intentionally left blank

Editorial Assistants and Reviewers for MIE2020

Editorial Assistant Acknowledgements

The editors would like to gratefully acknowledge the assistance of our colleagues in final preparation of the manuscripts for publication.

Adedeji, Taiwo

Nakkas, Haythem

Ozonze, Obinwa

Foufi, Vasiliki, Dr. PhD

Gaudet-Blavignac, Christophe, BSc CS, MMed

Reviewers

The editors would also like to gratefully thank the reviewers who have contributed in the reviewing phase. Without these reviews we would not be able to have so high standards.

This page intentionally left blank

Contents

Digital Personalized Health and Medicine: Preface	v
<i>Philip Scott, Louise B. Pape-Haugaard, Christian Lovis, Inge Cort Madsen, Patrick Weber, Per Hostrup Nielsen</i>	
Digital Personalized Health and Medicine: COVID-19	vii
<i>Louise B. Pape-Haugaard, Lăcrămioara Stoicu-Tivadar, Inge Cort Madsen, Alfred Winter and Christian Lovis</i>	
Editorial Assistants and Reviewers for MIE2020	ix

Part 1

Section 1. Biomedical Data, Tools and Methods

A FHIR-Based eConsent App for the Digital Hospital	3
<i>Manuela Weber, Alexandra Griessbach, Regina Grossmann and Jürg Blaser</i>	
A Secure Multi-Party Computation Protocol for Time-To-Event Analyses	8
<i>Lennart Vogelsang, Moritz Lehne, Phillipp Schoppmann, Fabian Prasser, Sylvia Thun, Björn Scheuermann and Josef Schepers</i>	
A Semantic Similarity Evaluation for Healthcare Ontologies Matching to HL7 FHIR Resources	13
<i>Athanasios Kiourtis, Argyro Mavrogiorgou and Dimosthenis Kyriazis</i>	
A Semi-Automated Approach for Multilingual Terminology Matching: Mapping the French Version of the ICD-10 to the ICD-10 CM	18
<i>Emmanuelle Sylvestre, Guillaume Bouzillé, Michael McDuffie, Emmanuel Chazard, Paul Avillach and Marc Cuggia</i>	
An Automatic Image Collection System for Multicenter Clinical Studies	23
<i>Toshihiro Takeda, Shirou Manabe, Atsushi Hattori, Junji Yamaguchi, Shozo Konishi, Yuichiro Yamamoto, Daiyo Takahashi and Yasushi Matsumura</i>	
An Evolutionary Approach to the Annotation of Discharge Summaries	28
<i>Christina Lohr, Luise Modersohn, Johannes Hellrich, Tobias Kolditz and Udo Hahn</i>	
Application of Radiomics in Vesselness Analysis of CT Angiography Images of Stroke Patients	33
<i>Daria Dolotova, Ivan Arkhipov, Evgenia Blagosklonova, Victoria Donitova, Tatiana Barmina, Faat Sharifullin, Boris Kobrinskii and Andrey Gavrilov</i>	
Automated Mapping of LEP Nursing Data to Nursing Minimum Data Sets	38
<i>Renate Ranegger, Werner O. Hackl, Inge Eberl, Dieter Baumberger, Reto Bürgin and Elske Ammenwerth</i>	

Automated Spelling Correction for Clinical Text Mining in Russian <i>Ksenia Balabaeva, Anastasia Funkner and Sergey Kovalchuk</i>	43
Automatic Classification of Discharge Letters to Detect Adverse Drug Reactions <i>Vasiliki Foufi, Kuntheavy Ing Lorenzini, Jean-Philippe Goldman, Christophe Gaudet-Blavignac, Christian Lovis and Caroline Samer</i>	48
Automatic Extraction of Risk Factors for Dialysis Patients from Clinical Notes Using Natural Language Processing Techniques <i>George Michalopoulos, Hammad Qazi, Alexander Wong, Zahid Butt and Helen Chen</i>	53
Automating Quality Control for Structured Standardized Radiology Reports Using Text Analysis <i>Anjani Dhrangadhariya, Sandy Millius, Cyril Thouly, Benoit Rizk, Dominique Fournier, Henning Müller and Hugues Brat</i>	58
Bank of Digital Filters for Identification of Combined Drug Products in a French Public Database <i>Adrien Ugon, Mélanie Courtine, Karima Sedki, Pierre Quentin, Marion Valette and Rosy Tsopra</i>	63
Better Safe than Sorry – Implementing Reliable Health Data Anonymization <i>Raffael Bild, Klaus A. Kuhn and Fabian Prasser</i>	68
Blood Lactate Concentration Prediction in Critical Care <i>Behrooz Mamandipoor, Mahshid Majd, Monica Moz and Venet Osmani</i>	73
Building an I2B2-Based Population Repository for Clinical Research <i>Lydia González, David Pérez-Rey, Enrique Alonso, Gema Hernández, Pablo Serrano, Miguel Pedrera, Agustín Gómez, Kristof De Schepper, Titus Crepain and Brecht Claerhout</i>	78
Character-Level Neural Language Modelling in the Clinical Domain <i>Markus Kreuzthaler, Michel Oleynik and Stefan Schulz</i>	83
Clinical Abbreviation Disambiguation Using Deep Contextualized Representation <i>Mingkai Peng and Hude Quan</i>	88
Clinical Concept Normalization on Medical Records Using Word Embeddings and Heuristics <i>João Figueira Silva, Rui Antunes, João Rafael Almeida and Sérgio Matos</i>	93
Clinical History Segment Extraction from Chronic Fatigue Syndrome Assessments to Model Disease Trajectories <i>Sonia Priou, Natalia Viani, Veshalee Vernugopan, Chloe Tytherleigh, Faduma Abdalla Hassan, Rina Dutta, Trudie Chalder and Sumithra Velupillai</i>	98
Clinico-Environmental System for Personal Monitoring <i>Filipe Pires, André Pedrosa, João Alegria and Carlos Costa</i>	103

Cohort Creation and Visualization Using Graph Model in the PREDIMED Health Data Warehouse	108
<i>Christophe Cancé, Pierre-Ephrem Madiot, Christian Lenne, Svetlana Artemova, Brigitte Cohard, Marjolaine Bodin, Alban Caporossi, Jean-François Blatier, Jérôme Fauconnier, Frédérique Olive, Daniel Pagonis, Dominique Le Magny, Jean-Luc Bosson, Katia Charriere, Ivan Paturel, Bruno Lavaire, Gabriel Schummer, Joseph Eterno, Jean-Noël Ravey, Ivan Bricault, Gilbert Ferretti, Sébastien Chanoine, Pierrick Bedouch, Emmanuel Barbier, Julien Thevenon, Pascal Mossuz and Alexandre Moreau-Gaudry</i>	
Commercial Adoption of AI in the Healthcare Sector: An Exploratory Analysis of S&P500 Companies	113
<i>Tomasz Mucha, Olli Halminen, Henni Tenhunen and Timo Seppälä</i>	
A Computational Adverse Event Detection Matrix	118
<i>Mary Gadde and Melody Penning</i>	
Computer-Assisted Quality Assessment of Aortic CT Angiographies for Patient-Individual Dose Adjustment	123
<i>Marja Fleitmann, Kira Soika, Andreas Martin Stroth, Jan Gerlach, Alexander Fürschke, Peter Hunold, Jörg Barkhausen, Arpad Bischof and Heinz Handels</i>	
Continuous Monitoring and Statistical Modelling of Heart Rate Variability	128
<i>Adina Nitulescu, Mihaela Crisan-Vida and Lacramioara Stoicu-Tivadar</i>	
Data Fusion to Convert Drug Consumption Quantities into Defined Daily Doses	133
<i>Adrien Ugon, Sabine Guessant and Rosy Tsopra</i>	
Data Integration into OMOP CDM for Heterogeneous Clinical Data Collections via HL7 FHIR Bundles and XSLT	138
<i>Patrick Fischer, Mark R. Stöhr, Henning Gall, Achim Michel-Backofen and Raphael W. Majeed</i>	
Data Quality Challenges in a Learning Health System	143
<i>Michail Sarafidis, Marilena Tarousi, Athanasios Anastasiou, Stavros Pitoglou, Efstratios Lampoukas, Athanasios Spetsarias, George Matsopoulos and Dimitrios Koutsouris</i>	
De-Identifying Swedish EHR Text Using Public Resources in the General Domain	148
<i>Taridzo Chomutare, Kassaye Yitbarek Yigzaw, Andrius Budrionis, Alexandra Makhlysheva, Fred Godtliebsen and Hercules Dalianis</i>	
Deep Learning for Accurate Diagnosis of Glaucomatous Optic Neuropathy Using Digital Fundus Image: A Meta-Analysis	153
<i>Mohaimenul Islam, Tahmina Nasrin Poly, Hsuan Chia Yang, Suleman Atique and Yu-Chuan Jack Li</i>	

Design for a Modular Clinical Trial Recruitment Support System Based on FHIR and OMOP	158
<i>Ines Reinecke, Christian Gulden, Michéle Kümmel, Azadeh Nassirian, Romina Blasini and Martin Sedlmayr</i>	
Detection of Muscle Weakness in Medical Texts Using Natural Language Processing	163
<i>Gleb Danilov, Michael Shifrin, Yuliya Strunina, Konstantin Kotik, Tatyana Tsukanova, Tatiana Pronkina, Timur Ishankulov, Elizaveta Makashova, Alexandra Kosyrkova and Alexander Potapov</i>	
Digitalisation of the Brief Visuospatial Memory Test-Revised and Evaluation with a Machine Learning Algorithm	168
<i>Martin Eduard Birchmeier, Tobias Studer, Andreas Lutterotti, Iris-Katharina Penner and Serge Bignens</i>	
Discovering New Social Determinants of Health Concepts from Unstructured Data: Framework and Evaluation	173
<i>Joao H. Bettencourt-Silva, Natalia Mulligan, Marco Sbodio, John Segrave-Daly, Richard Williams, Vanessa Lopez and Carlos Alzate</i>	
Distinguishing Septal Heart Defects from the Valvular Regurgitation Using Intelligent Phonocardiography	178
<i>Arash Gharehbaghi, Amir A. Sepehri and Ankica Babic</i>	
Document Oriented Graphical Analysis and Prediction	183
<i>Shorabuddin Syed, Mahanazuddin Syed, Hafsa Bareen Syeda, Fred Prior, Meredith Zozus, Melody L. Penning and Mohammed Orloff</i>	
Electronic Whiteboard in the Inpatient Care Management: Nurses' Perceptions About the Use	188
<i>Liliana Giraldo, Bibiana Schachner, Daniel Luna and Sonia Benitez</i>	
Efficient Protection of Health Data from Sensitive Attribute Disclosure	193
<i>Raffael Bild, Johanna Eicher and Fabian Prasser</i>	
Emerging Concepts and Applied Machine Learning Research in Patients with Drug-Induced Repolarization Disorders	198
<i>Mina Bjelogrljic, Arnaud Robert, Arnaud Miribel, Mehdi Namdar, Baris Gencer, Christian Lovis and François Girardin</i>	
End-to-End Approach for Structuring Radiology Reports	203
<i>Kento Sugimoto, Toshihiro Takeda, Shoya Wada, Asuka Yamahata, Shozo Konishi, Shiro Manabe and Yasushi Matsumura</i>	
Evaluation of Document Retrieval Systems on a Medical Corpus in French: Indexation vs. Feature Learning	208
<i>Arnaud Robert, Francis Damachi, Mina Bjelogrljic, Jean-Philippe Goldman and Christian Lovis</i>	
Evaluation of the Quality of French Hospital Data for Perinatal Algorithms	213
<i>Karine Goueslard, Jonathan Cottenet, Eric Benzenine, Pascale Tubert-Bitter and Catherine Quantin</i>	

Exploring Patient Path Through Sankey Diagram: A Proof of Concept <i>Antoine Lamer, Gery Laurent, Sylvia Pelayo, Mehdi El Amrani, Emmanuel Chazard and Romaric Marcilly</i>	218
Extending Contsys Standard with Social Care Concepts: A Methodology Proposed by the UNINFO Working Group in Italy <i>Fabrizio Pecoraro, Daniela Luzi, Elaheh Pourabbas, Fabrizio L. Ricci and Angelo Rossi Mori</i>	223
Feasibility of Using EN 13606 Clinical Archetypes for Defining Computable Phenotypes <i>Archana Tapuria, Dipak Kalra and Vasa Curcin</i>	228
Generation of Fine Grained Demographic Information for Epidemiological Analysis <i>Timo Wolters, Oke Wübbenhorst, Christian Lüpkes and Andreas Hein</i>	233
GenoShare: Supporting Privacy-Informed Decisions for Sharing Individual-Level Genetic Data <i>Jean Louis Raisaro, Juan Ramón Troncoso-Pastoriza, Yamane El-Zein, Mathias Humbert, Carmela Troncoso, Jacques Fellay and Jean-Pierre Hubaux</i>	238
H-Accuracy, an Alternative Metric to Assess Classification Models in Medicine <i>Andrea Campagner, Luca Sconfienza and Federico Cabitza</i>	242
Heimdall, a Computer Program for Electronic Health Records Data Visualization <i>Niels Martignene, Thibaut Balcaen, Guillaume Bouzille, Matthieu Calafiore, Jean-Baptiste Beuscart, Antoine Lamer, Bertrand Legrand, Grégoire Ficheur and Emmanuel Chazard</i>	247
How Have Experimental Cancer Interventions Evolved over Time? <i>Ziran Li, Huanyao Chen, Chi Yuan and Chunhua Weng</i>	252
Impact of Altering Data Granularity Levels on Predictive Modelling: A Case Study of Fall Risk Prediction in Older Persons <i>Noman Dormosh, Ameen Abu-Hanna, Nathalie van der Velde and Martijn Schut</i>	257
Incorporation of Multiple Sources into IT – and Data Protection Concepts: Lessons Learned from the FARKOR Project <i>Doris Lindoerfer, Ulrich Mansmann and Isabel Reinhardt</i>	262
Integrating the Comparative Toxicogenomic Database in a Human Pharmacogenomic Resource <i>François-Élie Calvier and Cedric Bousquet</i>	267
Integration of Unstructured Data into a Clinical Data Warehouse for Kidney Transplant Screening – Challenges & Solutions <i>Maximilian Zubke, Matthias Katzensteiner and Oliver J. Bott</i>	272

Interpretation Method for Continuous Glucose Monitoring with Subsequence Time-Series Clustering	277
<i>Masaki Ono, Takayuki Katsuki, Masaki Makino, Kyoichi Haida and Atsushi Suzuki</i>	
Introducing New Measures of Inter- and Intra-Rater Agreement to Assess the Reliability of Medical Ground Truth	282
<i>Andrea Campagner and Federico Cabitza</i>	
Learning Scan Paths of Eye Movement in Autism Spectrum Disorder	287
<i>Meenakshi Chatterjee, Nikolay V. Manyakov, Abigail Bangerter, Dzmitry A. Kaliukhovich, Shyla Jagannatha, Seth Ness and Gahan Pandina</i>	
Leveraging PubMed to Create a Specialty-Based Sense Inventory for Spanish Acronym Resolution	292
<i>Alexandra Pomares-Quimbaya, Pilar López-Úbeda, Michel Oleynik and Stefan Schulz</i>	
Logical Rules and a Preliminary Prototype for Translating Mortality Coding Rules from ICD-10 to ICD-11	297
<i>Vincenzo Della Mea, Mihai H. Popescu, Francesco Grippo, Chiara Orsi and Friedrich Heuser</i>	
Machine Learning Assisted Citation Screening for Systematic Reviews	302
<i>Anjani Dhrangadhariya, Roger Hilfiker, Roger Schaer and Henning Müller</i>	
Machine Learning Explainability in Breast Cancer Survival	307
<i>Tom Jansen, Gijs Geleijnse, Marissa Van Maaren, Mathijs P. Hendriks, Annette Ten Teije and Arturo Moncada-Torres</i>	
Machine Learning for Automatic Encoding of French Electronic Medical Records: Is More Data Better?	312
<i>Julien Gobeill, Patrick Ruch and Rodolphe Meyer</i>	
MedCo ² : Privacy-Preserving Cohort Exploration and Analysis	317
<i>David Froelicher, Mickaël Misbach, Juan R. Troncoso-Pastoriza, Jean Louis Raisaro and Jean-Pierre Hubaux</i>	
Mixed and Augmented Reality Tools in the Medical Anatomy Curriculum	322
<i>Melanie Romand, Daniel Dugas, Christophe Gaudet-Blavignac, Jessica Rochat and Christian Lovis</i>	
Modulation of Medical Condition Likelihood by Patient History Similarity	327
<i>Jonathan Turner, Dymrna O'Sullivan and Jon Bird</i>	
MQT-TZ: Secure MQTT Broker for Biomedical Signal Processing on the Edge	332
<i>Carlos Segarra, Ricard Delgado-Gonzalo and Valerio Schiavoni</i>	
Near Real Time EHR Data Utilization in a Clinical Study	337
<i>Melody L. Penning, Collette Blach, Anita Walden, Pei Wang, Katrina M. Donovan, Maryam Y. Garza, Zhan Wang, Julie Frund, Shorabuddin Syed, Mahanazuddin Syed, Guilherme Del Fiol, L. Kristin Newby, Carl Pieper and Meredith Zozus</i>	

Negation Detection for Clinical Text Mining in Russian <i>Anastasia Funkner, Ksenia Balabaeva and Sergey Kovalchuk</i>	342
On the Construction of Multilingual Corpora for Clinical Text Mining <i>Fabián Villena, Urs Eisenmann, Petra Knaup, Jocelyn Dunstan and Matthias Ganzinger</i>	347
Ontology-Based and Architecture-Based Method for the Development of Interoperable Care Systems for Type 2 Diabetes Mellitus <i>Gustavo A. Uribe, Bernd Blobel, Diego M. Lopez, Stefan Schulz and Alonso A. Ruiz</i>	352
Optimization of Entropy-Based Automated Dyslalia Screening Algorithm <i>Emilian Erman Mahmut, Dorin Berian, Michele Della Ventura and Vasile Stoicu Tivadar</i>	357
Parallel Sentence Alignment from Biomedical Comparable Corpora <i>Rémi Cardon and Natalia Grabar</i>	362
Performances of a Solution to Semi-Automatically Fill eCRF with Data from the Electronic Health Record: Protocol for a Prospective Individual Participant Data Meta-Analysis <i>Nicolas Griffon, Helena Pereira, Juliette Djadi-Prat, María Teresa García, Sara Testoni, Manon Cariou, Jacques Hilbey, Aurèle N'Dja, Grégory Navarro, Nicola Gentili, Oriana Nanni, Massimo Raineri, Gilles Chatellier, Agustín Gómez De La Camara, Martine Lewi, Mats Sundgren, Christel Daniel, Almenia Garvey, Marija Todorovic and Nadir Ammour</i>	367
Physician-Machine Interaction in the Decision Making Process <i>Saveli Goldberg, Anatoly Temkin and Benjamin Weisburd</i>	372
Predicting Length of Stay in Hospital Using Electronic Records Available at the Time of Admission <i>Marta Wilk, D. William R. Marsh, Sarah De Freitas and John Prowle</i>	377
Predicting Postoperative Hospital Stay in Neurosurgery with Recurrent Neural Networks Based on Operative Reports <i>Gleb Danilov, Konstantin Kotik, Michael Shifrin, Uliya Strunina, Tatyana Pronkina and Alexander Potapov</i>	382
Probabilistic Approaches to Overcome Content Heterogeneity in Data Integration: A Study Case in Systematic Lupus Erythematosus <i>Alexia Sampri, Nophar Geifman, Helen Le Sueur, Patrick Doherty, Philip Couch, Ian Bruce, Niels Peek and MASTERplans Consortium</i>	387
Problems in FAIRifying Medical Datasets <i>Matthias Löbe, Franz Matthies, Sebastian Stäubert, Frank A. Meineke and Alfred Winter</i>	392
Prototyping a Tool for Processing Genetic Meta-Data in Microbiological Laboratories <i>Jan Aeschimann, Silvan Huber, Daniel Wüthrich, Helena Seth-Smith, Jürgen Holm, Thomas Bürkle and Murat Sariyar</i>	397

Providing an Integrated Access to EHR Using <i>Electronic Health Records Aggregators</i>	402
<i>Belen Prados-Suarez, Carlos Molina and Carmen Peña-Yañez</i>	
Retrieving and Analyzing Hospital Service Suspensions from Regional Healthcare Insurance Claims Data	407
<i>Jumpei Sato, Kazutoshi Umemoto, Kazuo Goda, Masaru Kitsuregawa and Naohiro Mitsutake</i>	
Secure Collapsing Method Based on Fully Homomorphic Encryption	412
<i>David Niyitegeka, Reda Bellafqira, Emmanuelle Genin and Gouenou Coatrieux</i>	
Semantic Querying of Hospital Data Using an Ontology-Based Model of Discharge Summaries and ICD 10	417
<i>Ihssen Belhadj and Thierry Boudemaghe</i>	
Supervised Bayesian Statistical Learning to Identify Prognostic Risk Factor Patterns from Population Data	422
<i>Colin J. Crooks</i>	
Supervised Learning for the ICD-10 Coding of French Clinical Narratives	427
<i>Clément Dalloux, Vincent Claveau, Marc Cuggia, Guillaume Bouzillé and Natalia Grabar</i>	
The Impact of Specialized Corpora for Word Embeddings in Natural Language Understanding	432
<i>Antoine Neuraz, Bastien Rance, Nicolas Garcelon, Leonardo Campillos Llanos, Anita Burgun and Sophie Rosset</i>	
The LOINC Content Model and Its Limitations of Usage in the Laboratory Domain	437
<i>Cora Drenkhahn and Josef Ingenerf</i>	
The openEHR Genomics Project. To the memory of Gianluigi Zanetti	443
<i>Cecilia Mascia, Francesca Frexia, Paolo Uva, Gianluigi Zanetti, Luca Pireddu, Gideon Giacomelli, Christina Jaeger-Schmidt, Aurelie Tomczak, Simon Schumacher, Florian Kraecher, Roland Eils, Silje Ljosland Bakke and Heather Leslie</i>	
The Personal Health Library: A Single Point of Secure Access to Patient Digital Health Information	448
<i>Nariman Ammar, James E. Bailey, Robert L. Davis and Arash Shaban-Nejad</i>	
The SmartSight Project: Use of Electronic Glasses to Improve Impaired Fields of Vision	453
<i>Berglind F. Smaradottir, Niels F. Garmann-Johnsen, Steinar Omnes, Ann-Elisabeth Ludvigsen and Harald Reiso</i>	
Three-Dimensional Volumetric Renal Reconstruction Based on Geometrical Coefficients	458
<i>Tomasz Les, Tomasz Markiewicz, Mirosław Dziekiewicz and Malgorzata Lorent</i>	

Towards Precise Descriptions of Medical Free/Libre and Open Source Software <i>Franziska Jahn, Michelle Bindel, Konrad Höffner, Maryam Ghalandari, Birgit Schneider, Sebastian Stäubert, Verena Dornauer, Thomas Karopka, Elske Ammenwerth and Alfred Winter</i>	463
Trajectories of Disease Accumulation Using Electronic Health Records <i>Pere Planell-Morell, Madhavi Bajekal, Spiros Denaxas, Rosalind Raine and Daniel C. Alexander</i>	469
Translating Social Determinants of Health into Standardized Clinical Entities <i>Michael Watkins, Benjamin Viernes, Viet Nguyen, Leonardo Rojas Mezarina, Javier Silva Valencia and Damian Borbolla</i>	474
Trodden Lanes or New Paths: Ballisto- and Seismocardiography Till Now <i>Nico Jähne-Raden, Henrike Gütschleg and Michael Marscholke</i>	479
Understanding Petri Nets in Health Sciences Education: The Health Issue Network Perspective <i>Fabrizio L. Ricci, Fabrizio Consorti, Fabrizio Pecoraro, Daniela Luzi, Valentina Mingarelli, Sara Miotti and Oscar Tamburis</i>	484
Using Big Data Analytics to Identify Dentists with Frequent Future Malpractice Claims <i>Wanting Cui and Joseph Finkelstein</i>	489
Using Business Intelligence Tools to Support Medical Validation of Laboratory Tests <i>Gian-Andrea Degen, Viola Günther, Jürgen Holm, Thomas Bürkle and Murat Sariyar</i>	494
Using Unsupervised Learning to Identify Clinical Subtypes of Alzheimer's Disease in Electronic Health Records <i>Nonie Alexander, Daniel C. Alexander, Frederik Barkhof and Spiros Denaxas</i>	499
Assessment of the Prognostic Accuracy of Scores in Different Ways on the Example of Assessing the Validity of the Prognostic Model ISS-RTS-TRISS <i>Irina V. Vasilyeva, Tatyana V. Zarubina and Yuri D. Udalov</i>	504
VIVID: Independent Living of Aging Adults Suffered a Stroke <i>Athanasios Anastasiou, Ourania Manta, Dimitra Iliopoulou and Dimitrios Koutsouris</i>	509
Section 2. Supporting Care Delivery	
A Domain-Independent Semantically Validated Authoring Tool for Formalizing Clinical Practice Guidelines <i>Jordi Torres, Garazi Artola and Naiara Muro</i>	517
A Process Mining Application for the Analysis of Hospital-at-Home Admissions <i>Ilaria Angela Amantea, Emilio Sulis, Guido Boella, Renata Marinello, Dario Bianca, Enrico Brunetti, Mario Bo and Carlos Fernandez-Llatas</i>	522

An Albanian Text-to-Speech System for the BabelDr Medical Speech Translator <i>Nikos Tsourakis, Rovena Troqe, Johanna Gerlach, Pierrette Bouillon and Hervé Spechbach</i>	527
Are Clinical Decision Support Systems Compatible with Patient-Centred Care? <i>Vije Kumar Rajput, Jack Dowie and Mette Kjer Kaltoft</i>	532
Case-Based Reasoning for Support of the Diagnostics of Cardiovascular Diseases <i>Zuzana Tocimáková, Ludmila Pusztová, Ján Paralič and Dominik Pella</i>	537
Clinical Decision Support Systems in Nursing Homes: A Scoping Review <i>Abir Abdellatif, Jacques Bouaud, Duy Nghiem, Carmelo Lafuente-Lafuente, Joël Belmin and Brigitte Séroussi</i>	542
Comparison of Unplanned 30-Day Readmission Prediction Models, Based on Hospital Warehouse and Demographic Data <i>Thibault Dhalluin, Aurélie Bannay, Pierre Lemordant, Emmanuelle Sylvestre, Emmanuel Chazard, Marc Cuggia and Guillaume Bouzille</i>	547
Cross-Border Exchange of Clinical Data Using Archetype Concepts Compatible with the International Patient Summary <i>Dimitar Tcharaktchiev, Evgeniy Krastev, Patrick Petrossians, Simeon Abanos, Hristo Kyurkchiev and Petko Kovatchev</i>	552
Data Transfer Model for HIS and Developers Opinions in Turkey <i>Filiz Isleyen and M. Mahir Ulgu</i>	557
Decision Quality Is a Preference-Sensitive Formative Concept: How Do Some Existing Measures Compare? <i>Jack Dowie and Mette Kjer Kaltoft</i>	562
Defining the Contextual Problem List <i>John Meredith, Ian McNicoll, Nik Whitehead and Kemi Ademoye</i>	567
Digital Follow-Up Application for Cancer Patients – Value Mechanisms Regarding Health Professionals' Work <i>Sauli Miettinen and Henni Tenhunen</i>	572
Digital Technology Trends Supporting Assisted Independent Living of Ageing Population <i>Anthony Maeder</i>	577
Electronic Health Records as a Facilitator for the Implementation of a Strategy for the Reduction of Maternal and Child Morbi-Mortality in Buenos Aires City <i>Ivan Recalde, Denise Levi, Mariano Franco, Santiago Esteban and Analía Baum</i>	582
Evaluating Quality of Electronic Clinical Notes Using the Spanish Version of Physician Documentation Quality Instrument (PDQI-9) in a High-Complexity Hospital. Cross-Sectional Study <i>Giuliana Colussi, Amalia Botto, Romina Giudicessi, Laura Gomel, Aldo Marangone, Miguel Rodriguez and Sonia Benitez</i>	587

Evaluation of openEHR Repositories Regarding Standard Compliance <i>Maximilian Fechner, Tobias J. Brix, Tobias Hardt, Johannes Oehm, Martin Dugas and Michael Storck</i>	592
Formal Modelling of FHIR Based, Medical Data Exchange Using Algebraic Petri Nets <i>Franziska Bathelt, Michèle Kümmel, Sven Helfer, Christian Kamann and Martin Sedlmayr</i>	597
Health Metrics Network (HMN-WHO), a Tool to Assess the Quality of the Information System in Burkina Hospitals <i>Seydou Golo Barro, Amadou Dicko, Paul-André De Lame and Pascal Staccini</i>	602
Implementation and Analysis of Two Knowledge Base Approaches for the Treatment of Chronic Wounds <i>Stefan Vogel, Jendrik Richter, Stefanie Wache, Kerstin Pischek-Koch, Simon Auchter, Sebastian Zebbities, Karen Güttler, Ursula Hübner, Mareike Pryzsucha, Jens Hüsters and Björn Sellemann</i>	607
Information Capturing in Pre-Hospital Emergency Medical Settings (EMS) <i>Kerstin Denecke, Lea Meier, Jan G. Bauer, Matthias Bender and Christopher Lueg</i>	613
Integrating Pharmacogenetic Decision Support into a Clinical Information System <i>Kevin Tippenhauer, Marwin Philips, Carlo R. Largiadèr, Murat Sariyar and Thomas Bürkle</i>	618
Localisation, Personalisation and Delivery of Best Practice Guidelines on an Integrated Care and Cure Cloud Architecture: The C3-Cloud Approach to Managing Multimorbidity <i>George Despotou, Gokce B. Laleci Erturkmen, Mustafa Yuksel, Bunyamin Sarigul, Pontus Lindman, Marie-Christine Jaulent, Jacques Bouaud, Lamine Traore, Sarah N. Lim Choi Keung, Esteban De Manuel, Dolores Verdoy, Antonio De Blas, Nicolas Gonzalez, Mikael Lilja, Marie Sherman, Malte Von Tottleben, Marie Beach, Christopher Marguerie, Liran Karni, Gunnar O. Klein, Dipak Kalra, Rong Chen and Theodoros N. Arvanitis</i>	623
Making a Subjective Notion Computer-Interpretable: The Case of the Tumour-Volume to Breast-Volume Ratio for the Surgical Decision of Breast Cancer <i>Claudia Blancafort, Jacques Bouaud, Coralie Prebet and Brigitte Seroussi</i>	628
More Information Continuity Through Health Networks? Barriers to Implementation and Use of Health IT <i>Özlem Teckert</i>	633
Nursing Record Innovations Aimed at Harmonizing Structured Clinical Knowledge Among Doctors and Nurses <i>Satoko Tsuru, Tetsuro Tamamoto, Hitoshi Furuya, Akihiro Nakao, Mari Fukuyama, Kouichi Tanizaki and Naohisa Yahagi</i>	638

Physicians' Experiences of Patient-Initiated Online Consultations in Primary Care Using Direct-To-Consumer Technology <i>Sabine Koch and Miwa Guhres</i>	643
PM-TOM: A Method for Finding Personalized Polypharmacy Therapies with Minimal Adverse Drug-Drug, Drug-Gene and Drug-Condition Interactions <i>Adnan Kulenovic and Azra Lagumdzija-Kulenovic</i>	648
Practical Implementation of Receiver-Oriented Encryption in STROKE OWL <i>Timo Michelsen, Christian Lins, Andreas Hein and Christian Lüpkes</i>	653
Relationship Between Exercise Duration in Multimodal Telerehabilitation and Quality of Sleep in Patients with Multiple Sclerosis <i>In cheol Jeong, Herbert Karpatkin, Joel Stein and Joseph Finkelstein</i>	658
Role-Based Architecture for Secure Management of Telepathology Sessions <i>Rui Jesus, Pedro Nunes, Rui Lebre and Carlos Costa</i>	663
Safer Surgery Checklist: Barriers in the Adoption of Tablets in Operating Room <i>Giuliana Colussi, Gabriela Garcia, Maria Grande and Daniel Luna</i>	668
Sensor-Based Decision Support for the Allocation of Patient Attendants in Hospitals <i>Sascha Gfeller, Philip Kyburz, Kai-Uwe Schmitt, Thomas Bürkle, Jürgen Holm and Murat Sariyar</i>	673
Smart Ageing: Digital Solutions for Future Care <i>Anthony Maeder and Gary Morgan</i>	678
Statistically Prioritized and Contextualized Clinical Decision Support Systems, the Future of Adverse Drug Events Prevention? <i>Emmanuel Chazard, Jean-Baptiste Beuscart, Michaël Rochoy, Olivia Dalleur, Bertrand Decaudin, Pascal Odou and Grégoire Fichet</i>	683
Tackling Polypharmacy: A Multi-Source Decision Support System <i>Guilherme Redeker and Juliana Bowles</i>	688
The House of Carbs: Personalized Carbohydrate Dispenser for People with Diabetes <i>Pietro Randine, Daniela Micucci, Gunnar Hartvigsen and Eirik Årsand</i>	693
Update on the DMP, the French Nationally Shared Medical Record: Did We Make It? <i>Brigitte Séroussi and Jacques Bouaud</i>	698
User Requirements Meet Large-Scale EHR Suites: Norwegian Preparations for Epic <i>Gunnar Ellingsen and Morten Hertzum</i>	703
Value Mechanisms in the Implementation of Intelligent Patient Flow Management System – A Multiple Case Study <i>Tuomo Tanila, Henni Tenhunen and Petteri Hirvonen</i>	708
Virtual Reality Therapy for Social Phobia: A Scoping Review <i>Elham Salehi, Mahdi Mehrabi, Farhad Fatehi and Asiyeh Salehi</i>	713

Visualizing the Cascade Effect of Redesigning Features in an EMR System <i>Samar Helou, Victoria Abou-Khalil, Goshiro Yamamoto, Osamu Sugiyama and Tomohiro Kuroda</i>	718
Web Application for Home Care Providers <i>Virginia Sandulescu, Sorin Puscoci, Monica Petre, Alexandru Girlea and Viorel Bota</i>	723

Part 2

Section 3. Health and Prevention

A Model for Early Identification of Chronic Obstructive Pulmonary Disease <i>Thomas Kronborg, Stine Hangaard and Ole Hejlesen</i>	731
A Study of the Possibility of Detecting Pediatric Mild Developmental Delay Through a Serious Game: A Randomized Cluster Trial in Cambodia <i>Ho Heon Kim, Hong Wook Kwon, Sambath Kao, Hyuk-Sang Sohn, Joohye Kim, Yejin Choi, Doora Kim, Mirae Kang and Yu Rang Park</i>	736
Current Use of Sensor-Based Measurements for Paraplegics: A Literature Review <i>Nils-Hendrik Benning, Niclas Hagen and Petra Knaup</i>	741
Detection of Psychomotor Agitation Pattern from Motion Sensor Data in a Living Environment of a Patient with Dementia <i>Corinna Mielke, Rasmus Antons and Reinhold Haux</i>	746
ECG-Quality Assessment of Dry-Electrode Cooperative Sensors <i>Michael Rapin, Josias Wacker, Damien Ferrario, Etienne Haenni, Stephan Dasen and Olivier Chéhelat</i>	751
Gesture Classification for a Hand Controller Device Using Neural Networks <i>Stelian Nicola, Oana Sorina Chirila and Lacramioara Stoicu-Tivadar</i>	756
Natural Language Processing for Detecting Medication-Related Notes in Heart Failure Telehealth Patients <i>Alphons Eggerth, Karl Kreiner, Dieter Hayn, Bernhard Pfeifer, Gerhard Pölzl, Tim Egelseer-Bründl and Günter Schreier</i>	761
Patient-Generated Health Data in the Clinic <i>Nirusha Jeevanandan and Christian Nøhr</i>	766
Register-Based Research of Adverse Events Revealing Incomplete Records Threatening Patient Safety <i>Ulla-Mari Kinnunen, Eija Kivekäs, Sari Palojoiki and Kaija Saranto</i>	771
Teletriage Pilot Study (Strategy for Unscheduled Teleconsultations): Results, Patient Acceptance and Satisfaction <i>Andrés Santiago Frid, María Florencia Grande Ratti, Ana Pedretti, Marina Valinoti, Bernardo Martínez, Janine Sommer, Daniel Luna and Fernando Plazzotta</i>	776

The Renal Health Instagram: An Analysis of Comments <i>Geraldo Bezerra da Silva Junior, Marjan Askari, Diovana Ximenes Cavalcante Dourado, Juliana Gomes Ramalho de Oliveira and José Eurico de Vasconcelos Filho</i>	781
Use of Social Media Apps by Nurses for Professional Purposes in Catalonia: Cross-Sectional Study <i>Antonio Torrejon, Octavi Rodriguez-Blanco and Miguel A. Mayer</i>	786
Visual Representation of African Traditional Medicine Recipes Using Icons and a Formal Ontology, ontoMEDTRAD <i>Appoh Kouame, Konan Marcellin Brou, Moussa Lo and Jean Baptiste Lamy</i>	791
What Do We Know About the Use of Chatbots for Public Health? <i>Elia Gabarron, Dillys Larbi, Kerstin Denecke and Eirik Årsand</i>	796
Section 4. Precision Medicine and Public Health	
A Stroll Along the Erroneous ePrescription Interactions Within the Belgian Pharmacy <i>Sven Van Laere, Ronald Buyl and Marc Nyssen</i>	803
A Structured Measurement of Highly Synchronous Real-Time Ballistocardiography Signal Data of Heart Failure Patients <i>Nico Jähne-Raden, Udo Bavendiek, Henrike Gütschleg, Ulf Kulau, Stephan Sigg, Marie Wolf, Tanja Zeppernick and Michael Marschollek</i>	808
Addressing the Search Challenges of Precision Medicine with Information Retrieval Systems and Physician Readers <i>Kate Fultz Hollis, Kirk Roberts, Steven Bedrick and William R. Hersh</i>	813
Argentinian Digital Health Strategy <i>Daniel A. Rizzato Lede, Federico A. Pedernera, Emiliano López, Cintia D. Speranza, Carlos Guevel, Jesse J. Maid, Patricia Mac Culloch, Floencia Rolandi, Fabiana Ayala, Daniel A. Abadie, María I. Baqué, Fernando Gassino, Fernando Campos, Diego Kaminker, Cintia A. Cejas, Alejandro López Osornio and Adolfo Rubinstein</i>	818
Assessing Opioid Use Patient Representations and Subtypes <i>Sarah Mullin and Peter Elkin</i>	823
Automated Engineering for Health Smart Homes: Find a Way in the Jungle of Assistance Systems <i>Bastian Wollschlaeger and Klaus Kabitzsch</i>	828
Decreasing Number of Medication Prescriptions After e-Prescriptions Became Mandatory and Their Valid Period Was Extended: A Big Bang Policy Change in Finland in 2017 <i>Vesa Jormanainen, Toni Relander, Vilma Jormanainen and Marina Lindgren</i>	833

Exploratory Analysis of HIV Status Knowledge and Associated Factors Using Data from Electronic Medical Records	838
<i>Natividad Burdisso, Santiago Esteban, Karin S. Kopitowski and Sergio A. Terrasa</i>	
Global Public Health Surveillance Using Media Reports: Redesigning GPHIN	843
<i>Dave Carter, Marta Stojanovic, Philip Hachey, Kevin Fournier, Simon Rodier, Yunli Wang and Berry De Bruijn</i>	
ICT Toolkit for Integrated Prevention, Testing and Linkage to Care Across HIV, Hepatitis, STIs and Tuberculosis in Europe	848
<i>Christine Kakalou, Lella Cosmaro, Vlasios K. Dimitriadis, Anne Raahauge, Dorthe Raben and Vassilis Koutkias</i>	
Improving Dengue Surveillance System with Administrative Claim Data in Indonesia: Opportunities and Challenges	853
<i>Atina Husnayain, Anis Fuad, Ida Safitri Laksono and Emily Chia-Yu Su</i>	
Intelligent Tools for Precision Public Health	858
<i>Anya Okhmatovskaia and David L. Buckeridge</i>	
Machine Learning-Based Identification of Obesity from Positive and Unlabelled Electronic Health Records	864
<i>Vicent Blanes-Selva, Salvador Tortajada, Ruth Vilar, Bernardo Valdivieso and Juan M. García-Gómez</i>	
Medical Informatics in the Digital Personalized Health and Medicine Era: A SWOT Analysis and Actionable Strategies	869
<i>Rada Hussein</i>	
Mining Potential Effects of HUMIRA in Twitter Posts Through Relational Similarity	874
<i>Keyuan Jiang, Shichao Feng, Liyuan Huang, Tingyu Chen and Gordon R. Bernard</i>	
Ontology-Guided Policy Information Extraction for Healthcare Fraud Detection	879
<i>Theodora S. Brisimi, Vanessa Lopez, Valentina Rho, Marco Sbodio, Gabriele Picco, Morten Kristiansen, John Segrave-Daly and Conor Cullen</i>	
Text-Mining Services of the Swiss Variant Interpretation Platform for Oncology	884
<i>Déborah Caucheteur, Julien Gobeill, Anaïs Mottaz, Emilie Pasche, Pierre-André Michel, Luc Mottin, Daniel J. Stekhoven, Valérie Barbié and Patrick Ruch</i>	
The Smart Device System for Movement Disorders: Preliminary Evaluation of Diagnostic Accuracy in a Prospective Study	889
<i>Julian Varghese, Michael Fajarski, Tim Hahn, Martin Dugas and Tobias Warnecke</i>	
User Expectations and Willingness to Share Self-Collected Health Data	894
<i>Ashenafi Zebene Woldaregay, André Henriksen, David-Zacharie Issom, Gerit Pfuhl, Keiichi Sato, Aude Richard, Christian Lovis, Eirik Årsand, Jessica Rochat and Gunnar Hartvigsen</i>	

Section 5. Human Factors and Citizen Centered Digital Health

A Path Towards Inclusion: Transdisciplinary Experience for the Inclusion of Self-Perceived Gender in an Information System	901
<i>Melanie Cassarino, Eunice Correa, Sebastián Minoletti, Amalia Botto, Romina Rapisarda, María Florencia Grande Ratti and Daniel Luna</i>	
A Systematic Review on CDSS Alert Appropriateness	906
<i>Olufisayo Olakotan, Maryati Mohd Yusof and Sharifa Ezat Wan Puteh</i>	
Analysing Sentiment and Topics Related to Multiple Sclerosis on Twitter	911
<i>Guido Giunti, Maëlick Claes, Enrique Dorronzoro Zubiete, Octavio Rivera-Romero and Elia Gabarron</i>	
Automatic Interpretation of Laboratory Tests and Its Influence on Follow-up	916
<i>Georgy Kopanitsa</i>	
Clinical-Based and Expert Selection of Terms Related to Depression for Twitter Streaming and Language Analysis	921
<i>Angela Leis, Miguel-Angel Mayer, Francesco Ronzano, Marta Torrens, Claudio Castillo, Laura I. Furlong and Ferran Sanz</i>	
Design of the CAPABLE Health Empowerment Tool: Citizens' Needs and Expectations	926
<i>Astri Letnes Janson, Anne Moen and Kristin Skeide Fuglerud</i>	
Designing a Patient-Centered Notification System into a PHR	931
<i>Mariana Simón, Giuliana Colussi, Janine Sommer, Lucila Bruchanski and Fernando Plazzotta</i>	
Developing an Intervention to Implement Electronic Patient-Reported Outcomes in Renal Services in the UK	936
<i>Sabine N. Van Der Veer, Angelo Ercia, Fergus J. Caskey, Ken Farrington, Francine Jury, Michael Rees, Tim Whitlock and Sarah Knowles</i>	
Development of a ePRO-Based Palliative Care Intervention for Cancer Patients: A Participatory Design Approach	941
<i>Christina Karamanidou, Christos Maramis, Kostas Stamatopoulos and Vassilis Koutkias</i>	
Development of a Mobile Digital Manikin to Measure Pain Location and Intensity	946
<i>Sabine N. Van Der Veer, Anna L. Beukenhorst, S. Mustafa Ali, Ben James, Pedro Silva, John McBeth and William G. Dixon</i>	
Digital Support for Self-Management in Children with Diabetes: Understanding Their Needs and Developing a Design Concept	951
<i>Nicole Safavi Andersen, Louise Hedrich Haugaard, Sissel Bøgh Pedersen, Mikkel Strange Pedersen and Ann Bygholm</i>	
Does Health-IT Improve People Centered Care?	956
<i>Pernille Bertelsen, Lone Stub Petersen and Christian Nøhr</i>	

eSource-Enabled vs. Traditional Clinical Trial Data Collection Methods: A Site-Level Economic Analysis <i>Eric L. Eisenstein, Maryam Y. Garza, Mitra Rocca, Gideon S. Gordon and Meredith Zozus</i>	961
Health Professionals' Experiences of the Benefits and Challenges of Online Symptom Checkers <i>Sari Kujala, Iris Hörhammer, Riitta Hänninen-Ervasti and Tarja Heponiemi</i>	966
How Come Nothing Changed? Reflections on the Fasting-Time Project <i>Morten Hertzum and Jesper Simonsen</i>	971
How to Evaluate Health Applications with Conversational User Interface? <i>Kerstin Denecke and Jim Warren</i>	976
Implementation of an Open-Source Electronic Health Record for Decision-Support Education in Medical Informatics <i>Stephanie Medlock, Roland Schermer and Ronald Cornet</i>	981
Implementation of EXABO – An Expert Advisory Board for the European Reference Network for Rare Respiratory Diseases <i>Désirée Walther, Olivia Steinmann, Thomas O.F. Wagner and Holger Storf</i>	986
Integrating a Secure and Generic Mobile App for Patient Reported Outcome Acquisition into an EHR Infrastructure Based on FHIR Resources <i>Nico Ploner and Hans-Ulrich Prokosch</i>	991
Interaction with Medical Venues in Megacities <i>Maria Óskarsdóttir and Anna Sigríður Íslind</i>	996
Investigating the Barriers to Physician Adoption of an Artificial Intelligence-Based Decision Support System in Emergency Care: An Interpretative Qualitative Study <i>Cécile Petitgand, Aude Motulsky, Jean-Louis Denis and Catherine Régis</i>	1001
Learning Interpretable Behavioral Engagement for Care Management <i>Subhro Das, Chandramouli Maduri, Ching-Hua Chen and Pei-Yun S. Hsueh</i>	1006
<i>Mi Argentina/Mi Salud</i> : The Argentinian Citizen Digital Health Portal <i>Daniel A. Rizzato Lede, Federico A. Pedernera, Emiliano López, Cintia D. Speranza, Jesse J. Maid, Fernando Gassino, Osvaldo Argibay, Rodrigo Castiglione, Guillermo Pujadas, Emanuel Gonzalez Alvaredo, Rodrigo Álvarez, Maria B. Markiewicz, Florencia Alegre, Leandro Ferrareis, Florencia Rolandi, Fabiana Ayala, Daniel A. Abadie, Cintia A. Cejas, Alejandro López Osornio and Adolfo Rubinstein</i>	1011
Nutrikids, a Smartphone Application to Improve the Quality of Paediatric Dietary Assessments: Feasibility Study <i>Léa Jacques, Catherine Bussien, Camille Descloux, Lisa Decorsaire, Christian Lovis and Frédéric Ehrler</i>	1016

On Patient Accessible Electronic Health Records and the Experienced Effect on the Work Environment of Nurses <i>Jonas Moll and Åsa Cajander</i>	1021
Online Medication Information for Citizens: A Comparison of Demands on eHealth Literacy <i>Helen Monkman, Thomas Schmidt and Christian Nøhr</i>	1026
Personalized Communication – People with Hearing Loss’ Experiences of Using a National Patient Portal <i>Maria Hägglund, Emma Helsing and Isabella Scandurra</i>	1031
Process Approach for Managing Health Information System-Induced Medication Errors <i>Maryati Mohd. Yusof, Toshihiro Takeda, Naoki Mihara and Yasushi Matsumura</i>	1036
Qualitative Evaluations of mHealth Interventions: Current Gaps and Future Directions <i>Meghan Bradway, Kari Leibowitz, Kathleen A. Garrison, Lauren Howe and Eirik Årsand</i>	1041
Social Robots for Elderly Care: An Inventory of Promising Use Cases and Business Models <i>Somaya Ben Allouch and Lex van Velsen</i>	1046
The Digital Transformation of Patient-Reported Outcomes’ (PROs) Functionality Within Healthcare <i>Jepe Eriksen, Pernille Bertelsen and Ann Bygholm</i>	1051
The Effect of Patient Accessible Electronic Health Records on Communication and Involvement in Care – A National Patient Survey in Sweden <i>Jonas Moll and Hanife Rexhepi</i>	1056
The Patient as Genomic Data Manager – Evaluation of the PROMISE App <i>Lena Griebel, Marc Hinderer, Ali Amr, Benjamin Meder, Marc Schweig, Dominic Deuber, Christoph Egger, Claudia Kawohl, Annika Krämer, Isabell Flade, Dominique Schröder and Hans-Ulrich Prokosch</i>	1061
Towards a Highly Usable, Mobile Electronic Platform for Patient Recruitment and Consent Management <i>Daniel Robins, Rachel Brody, In Cheol Jeong, Irena Parvanova, Jiazhen Liu and Joseph Finkelstein</i>	1066
Towards a Lean Process for Patient Journey Mapping – A Case Study in a Large Academic Setting <i>Marieke Sijm-Eeken, Jaqueline Zheng and Linda Peute</i>	1071
User Experience Theories, Models, and Frameworks: A Focused Review of the Healthcare Literature <i>Blake Lesselroth, Helen Monkman, Kathleen Adams, Scott Wood, Audrey Corbett, Juell Homco, Elizabeth M. Borycki, Ross Spier and Andre W. Kushniruk</i>	1076

User-Centred Approach to Design an Online Social Support Platform for Seniors: Identification of Users' Types and Their Requirements	1081
<i>Arnaud Ricci, Jessica Rochat, Henk Herman Nap, Lotte Cornelisse, Christian Lovis and Frédéric Ehrler</i>	

Section 6. Ethics, Legal and Societal Aspects

Autonomous Systems and Artificial Intelligence in Healthcare Transformation to 5P Medicine – Ethical Challenges	1089
<i>Bernd Blobel, Pekka Ruotsalainen, Mathias Brochhausen, Frank Oemig and Gustavo A. Uribe</i>	
Big Data, Information Technology and Information Professionals: Some Considerations for Digital Ethics	1094
<i>E-H. Kluge</i>	
Citizens' Participation in Health and Scientific Research in Switzerland	1098
<i>Jessica Rochat, Christophe Gaudet-Blavignac, Marzia Del Zotto, Victor Ruiz Garretas, Vasiliki Foufi, David Issom, Caroline Samer, Samia Hurst and Christian Lovis</i>	
Designing a Solution to Manage Electronic Consent for Children	1103
<i>Gary Leeming, Sarah Thew and John Ainsworth</i>	
Effectiveness of Web Accessibility Policy Implementation in Online Healthcare Information	1108
<i>Muzawir Arief, Sari Rissanen and Kaija Saranto</i>	
Evolution of Interdisciplinarity in Medical Informatics in Europe: Patterns from Intertwining Histories	1113
<i>Casimir A. Kulikowski and George I. Mihalas</i>	
General Data Protection Regulation (GDPR) in Healthcare: Hot Topics and Research Fronts	1118
<i>Farhad Fatehi, Farkhondeh Hassandoust, Ryan K.L. Ko and Saeed Akhlaghpour</i>	
Health 'Big Data' Value, Benefit, and Control: The Patient eHealth Equity Gap	1123
<i>Paul Demuro, Carolyn Petersen and Paul Turner</i>	
Human-Induced Errors in Networked Healthcare Research: Risk Management Under the GDPR	1128
<i>Jonas Bienenzeisler, Hauke Fischer, Volker S. Thiemann and Rainer Röhrig</i>	
Impact Analysis of the Policy for Access of Administrative Data in France: A Before-After Study	1133
<i>Vincent Looten and Marie Simon</i>	
Legal and Ethical Issues in Secondary Use of Administrative Health Data: The Case of Latvian Healthcare Monitoring Datalink	1138
<i>Signe Mezinska, Arnis Buka, Agnese Bankava and Juris Barzdins</i>	

The Definition of Informatics Competencies in Finnish Healthcare and Social Welfare Education	1143
<i>Alpo Värri, Minna Tiainen, Elina Rajalahti, Ulla-Mari Kinnunen, Lea Saarni and Outi Ahonen</i>	

Section 7. Swiss Personalized Health Network

SPHN – Development and Usability Testing of an Electronic General Consent Prototype	1151
<i>Julia Maurer, Ramon Saccilotto, Jörg Willers, Pascal Düblin, Vu Duong and Christiane Pauli-Magnus</i>	
SPHN – The Swiss Personalized Health Network Initiative	1156
<i>Adrien K. Lawrence, Liselotte Selter and Urs Frey</i>	
SPHN/PHRT – MedCo in Action: Empowering the Swiss Molecular Tumor Board with Privacy-Preserving and Real-Time Patient Discovery	1161
<i>Jean Louis Raisaro, Juan Ramón Troncoso-Pastoriza, Sylvain Pradervand, Michel Cuendet, Mickael Misbach, Joao Sa, Francesco Marino, Nicolas Freundler, Nicolas Rosat, David Cavin, Alexander Leichtle, Jacques Fellay, Olivier Michielin and Jean-Pierre Hubaux</i>	
SPHN/PHRT: Forming a Swiss-Wide Infrastructure for Data-Driven Sepsis Research	1163
<i>Adrian Egli, Manuel Battegay, Andrea C. Büchler, Peter Bühlmann, Thierry Calandra, Philippe Eckert, Hansjakob Furrer, Gilbert Greub, Stephan M. Jakob, Laurent Kaiser, Stephen L. Leib, Stephan Marsch, Nicolai Meinshausen, Jean-Luc Pagani, Jerome Pugin, Gunnar Rättsch, Jacques Schrenzel, Reto Schüpbach, Martin Siegemund, Nicola Zamboni, Reinhard Zbinden, Annelies Zinkernagel and Karsten Borgwardt</i>	
SPHN – The Swiss Aging Citizen Reference (SACR)	1168
<i>Ayoung Jeong, Murielle Bochud, Philippe Cattin, Manolis Dermitzakis, Bogdan Draganski, Andreas Papassotiropoulos, Martine Preisig, Bram Stieltjes, Peter Vollenweider and Nicole Probst-Hensch</i>	
SPHN – The BioMedIT Network: A Secure IT Platform for Research with Sensitive Human Data	1170
<i>Diana Coman Schmid, Katrin Crameri, Sabine Oesterle, Bernd Rinn, Thierry Sengstag, Heinz Stockinger, on behalf of the BioMedIT network team: Leila T. Alexander, Jonathan Barda, Christian Bolliger, Urban Borstnik, Gerhard Bräunlich, Olivier Byrde, Jérôme Dauvillier, Robin Engler, Pablo Escobar Lopez, Volker Flegel, Martin Fox, Sofia Georgakopoulou, Jani Heikkinen, Martin Jacquot, Nicolas Kowenski, Guillermo Losilla, Sergio Maffioletti, Jorge Molina, Diego Moreno, Allen Neeser, Michal Okoniewski, Warren Paulus, Kevin Sayers, Torsten Schwede, Jaroslaw Surkont, Richard Wartenburger and Thomas Wüst</i>	

Section 8. Posters

- A Digital Service Logistics Information System for Emergency Department
Care Coordination – Professionals’ Experiences 1177
*Laura-Maria Peltonen, Armi Kuloheimo, Kristiina Junttila
and Sanna Salanterä*
- A Matlab Tool for Organizing and Analyzing NHANES Data 1179
*Simon Lebech Cichosz, Morten Hasselstrøm Jensen,
Thomas Kronborg Larsen and Ole Hejlesen*
- A Path to Inclusion: Design and Prototype of Transgender Identity
in an Electronic Health Record 1181
*Melanie Cassarino, Eunice Correa, Sebastián Minoletti, Ignacio Jauregui,
Maria Florencia Grande Ratti, Carlos Otero and Daniel Luna*
- A Recommender System Based on Cohorts’ Similarity 1183
*João Rafael Almeida, Eriksson Monteiro, Luís Bastião Silva,
Alejandro Pazos and José Luís Oliveira*
- A Smart Mapping Editor for Standardised Data Transformation 1185
*Hannes Ulrich, Sebastian Germer, Ann-Kristin Kock-Schoppenhauer,
Jori Kern, Martin Lablans and Josef Ingenerf*
- Accessible Rates to Health Information on the Internet in Elderlies Increased
Among Fifteen Years 1187
Takeo Shibata, Akemi Kunitatsu, Chika Hiraga and Yoko Shimizu
- Addressing Ethics in the CrowdHEALTH RDI Project Concerned with Large
Amounts of Data to Support Health Policies 1189
Lydia Montandon
- Advanced Machine Learning in Prediction of Second Primary Cancer
in Colorectal Cancer 1191
Chi-Chang Chang and Ying-Chen Chen
- Agent-Based Modeling and Simulation of Care Delivery for Patients
with Thrombotic and Bleeding Disorders 1193
*Noemi Giordano, Samanta Rosati, Federica Valeri,
Alessandra Borchellini and Gabriella Balestra*
- An “in silico” Bench to Bedside Approach to Investigating Sepsis Biomarkers 1195
*Grace O’Mahony, Christopher Hawthorne, Karl McQuaid
and Guillermo H. Lopez-Campos*
- An Online Prevention Tool for Ophthalmological Disorders 1197
Maria Patte, Joseph Liaskos, Paris Gallos and John Mantas
- Analysis of Professional Competencies for the Clinical Research Data
Management Profession 1199
*Saly Abouelenein, Tremaine Williams, Jaime Baldner
and Meredith Nahm Zozus*

Application of IBM Watson to Support Literature Reviews: A Preliminary Experience in Restorative Dentistry	1201
<i>Giacomo Derchi, Marco Visentin, Vincenzo Marchio, Lisa Lardani, Antonio Barone, Marco Prenassi and Sara Marceglia</i>	
The Impact of Information Quality on Retracted Bioinformatics Literature	1203
<i>Austin Springer, Donna Chachere, Awaad Alsarkhi and Meredith Zozus</i>	
Association of C>U RNA Editing with Human Disease Variants	1205
<i>Sapan Mandloi, Zackary Falls, Rong Deng, Ram Samudrala and Peter L. Elkin</i>	
Automated Analysis of the Heterogeneity of Histological Glioblastoma Slides Using Neural Networks	1207
<i>Michael Oertl, Georg Prokop, Felix Holl, Marina Fotteler, Viktoria Muehlbauer, Walter Swoboda and Friederike Liesche-Starnecker</i>	
Automated Nerve Fibres Identification and Morphometry Analysis with Neural Network Based Tool in MATLAB	1209
<i>Michał Kopka, Wiktor Paskal, Adriana M. Paskal, Piotr Pietruski, Ryszard Kopka and Paweł K. Włodarski</i>	
Automatic Detection of Vital Signs in Clinical Notes of the Outpatient Settings	1211
<i>Martín Díaz Maffini, Fernanda Aguirre Ojea and Matías Manzotti</i>	
Best Practices for Designing Discrete Choice Experiments and the Use for Older Adults with Cognitive Impairment	1213
<i>Thomas Engelsma, Monique W.M. Jaspers and Anne M. Turner</i>	
Biostatistics Disruptive Acculturation Through Serious Gaming: A New Hope	1215
<i>Pierre Gillois, Lionel Di Marco, Josep M. Nicolàs, Alexandre Moreau-Gaudry, Anne Ego, Sandra David-Tchouda, Donald K. Martin and Jean-Luc Bosson</i>	
Challenges of In-House Development and Implementation of a CPOE for Oncology	1217
<i>S. de Matos Lima, C.L. Gimenez and D. Luna</i>	
Classification of Biceps Brachii Muscle Fatigue Condition Using Phase Space Network Features	1219
<i>Navaneethakrishna Makaram and Ramakrishnan Swaminathan</i>	
Classification of First-Episode Schizophrenia Using Wavelet Imaging Features	1221
<i>Kateřina Maršálová, Daniel Schwarz and Ivo Provazník</i>	
Classification of Histologic Images Using a Single Staining: Experiments with Deep Learning on Deconvolved Images	1223
<i>Vincenzo Della Mea and David Pilutti</i>	
Classification of Normal and Cardiomegaly Conditions in Chest Radiographs Using Cardio-Mediastinal Features	1225
<i>Satyavrat Govindarajan, Sukanta Kumar Tulo and Ramakrishnan Swaminathan</i>	

Classification of the Severity of Adverse Drugs Reactions <i>Raphaël Chauvet, Cédric Bousquet, Agnès Lillo-Lelouet, Ilan Zana, Ilan Ben Kimoun and Marie-Christine Jaulent</i>	1227
Coping with IT Downtime in Hospitals <i>Jackline Bamdeg and Thomas Schmidt</i>	1229
Creating and Implementing a Professional Development Program to Improve Organizational Capacity for Health Analytics in a Public Health Setting <i>Mariano Aizpurua, Tomás Galluzzi and Santiago Esteban</i>	1231
Cross-Mapping Study of Nursing Practice Terms from a Brazilian Hospital Database <i>Luciana Schleder Gonçalves, Dayane Andreatta and Fernanda Karoline Schamne</i>	1233
Data Integration Approaches for Representing Stem Cell Studies <i>Irena Parvanova and Joseph Finkelstein</i>	1235
Data Provenance Standards and Recommendations for FAIR Data <i>Malte-Levin Jauer and Thomas M. Deserno</i>	1237
Data Visualisation in Midwifery: The Challenge of Seeing what Datasets Hide <i>Bridget J. Daley, Evangelia Kyrimi, Kudakwashe Dube, Norman E. Fenton, Graham A. Hitman and Scott McLachlan</i>	1239
Deep Learning Approach for the Development of a Novel Predictive Model for Prostate Cancer <i>Mohaimenul Islam, Hsuan-Chia Yang, Phung-Anh Nguyen, Yu-Hsiang Wang, Tahmina Nasrin Poly and Yu-Chuan (Jack) Li</i>	1241
Defining a Standardized Information Model for Multi-Source Representation of Breast Cancer Data <i>Miguel Pedrera, Pablo Serrano, Ana Terriza, Jaime Cruz, Carolina Varela, Ana Salamanca, Noelia García, Carmen Salgado, Consuelo Sanz, Lucía Hernández, Montse Alemany and Adolfo Muñoz</i>	1243
Design of a Supportive Transfer Robot System for Caregivers to Reduce Physical Strain During Nursing Activities <i>Christian Kowalski, Pedro Arizpe-Gomez, Conrad Fifelski, Anna Brinkmann and Andreas Hein</i>	1245
Detecting Severe Incidents from Electronic Medical Records Using Machine Learning Methods <i>Kazuya Okamoto, Takashi Yamamoto, Shusuke Hiragi, Shosuke Ohtera, Osamu Sugiyama, Goshiro Yamamoto, Masahiro Hirose and Tomohiro Kuroda</i>	1247
Developing a Mobile App for Patients to Monitor Medical Record Changes Using Blockchain <i>SungJun Park, MinDong Sung, Sungjae Jung, Eunsol Lee, Jaehoon Lee and Yu Rang Park</i>	1249

Developing and Piloting TB Medication and Refilling Reminder System in Ethiopia <i>Kassahun Dessie Gashu, Fedlu Nurhussien, Adane Mamuye, Kassahun Alemu Gelaye and Binyam Tilahun</i>	1251
Development of a Graph-Based Database for Ovarian Cancer Symptoms <i>Young Ji Lee and Harleigh Niyu</i>	1253
Development of a Personalized mHealth Program for Cancer Symptom and Complication Management <i>Lixin Song, Eno Idiagbonya, Kevin Su and Juliet Idiagbonya</i>	1255
Development of a Questionnaire to Measure Healthcare Providers' Satisfaction with Telemedicine Services <i>Esmée Tensen, Youri Broekhuizen, Johan van Buggenum, Leonard Witkamp, Monique Jaspers and Linda Peute</i>	1257
Development of a Smart e-Coach Recommendation System for Obesity <i>Ayan Chatterjee, Martin W. Gerdes and Santiago Martinez</i>	1259
Development of a Systematic Text Annotation Standard to Extract Social Support Information from Electronic Medical Records <i>Camila Volij and Santiago Esteban</i>	1261
Digital Health Interventions for Diabetes Self-Management Education/Support in Type 1 & 2 Diabetes Mellitus <i>Dumisani Nkhom, Charles Jenya Soko, Pierre Bowrin and Usman Iqbal</i>	1263
Electronic Medical Records Virtual Course. Strategies for an Effective Training <i>Daniel A. Rizzato Lede, Cintia D. Speranza, Florencia Alegre, Adriana Cantora, Emilio Panizza and Ariadna Pou</i>	1265
Electronic Medication Administration System Supports Safe Medication Administration <i>Jenni Santavirta, Anne Kuusisto, Kaija Saranto, Tarja Suominen and Paula Asikainen</i>	1267
Emotion Analysis Using Electrodermal Signals and Spiking Deep Belief Network <i>Nagarajan Ganapathy and Ramakrishnan Swaminathan</i>	1269
Evaluating the Use of a CPOE for Chemotherapy Protocols <i>C. Gimenez, S. de Matos Lima and D. Luna</i>	1271
Evaluation of Mobile Phone Mortality Risk Score Applications Using Data from the Electronic Medical Records <i>Nino Fijacko, Lucija Gosak, Primož Koccek, Leona Cilar, Andrej Markota and Gregor Štiglic</i>	1273
EVENT: A Concept Developed to Improve the Use of EHR by the Nurses <i>Alain Junger</i>	1275

Experience and Expertise of Teledermatologists with Teledermoscopy: A Systematic Review to Guide Future Practice	1277
<i>Femke van Sinderen, Esmée Tensen, Job van der Heijden, Leonard Witkamp, Monique Jaspers and Linda Peute</i>	
Experience Story: How Do We Re-Implement What Has Been Implemented?	1279
<i>Melanie Cassarino, Camila Galvan, Juan Descalzo, Elio Jerez, María Smith and Daniel Luna</i>	
Experiences of National Documentation in Electronic Health Records: The Study Among Occupational Health Nurses	1281
<i>Sari Nissinen</i>	
Exploratory Analysis of Animal Bites Events in the City of Buenos Aires Using Data from Electronic Health Records	1283
<i>Gastón Nicolás Quintana and Santiago Esteban</i>	
Exploratory Analysis of Consultation Patterns Prior to the Diagnosis of Depression	1285
<i>Sofia Sciarreta and Santiago Esteban</i>	
Exploratory Analysis of Exercise Mentions in Clinical Notes	1287
<i>Nahuel Orcaizaguirre and Santiago Esteban</i>	
Feature Set for a Prediction Model of Diabetic Kidney Disease Progression	1289
<i>Masaki Ono, Takayuki Katsuki, Masaki Makino, Kyoichi Haida, Atsushi Suzuki and Reitaro Tokumasu</i>	
FHIR Driven Self-Management Support System for Diabetes	1291
<i>Dietmar Glachs, Tuncay Namli, Oliver Jung, Felix Strohmeier, Manuela Ploessnig and Gustavo Rodriguez</i>	
GDPR Compliant Blockchain and Distributed Ledger Technologies in the Health Sector	1293
<i>Anton Hasselgren, Katina Kravetska, Danilo Gligoroski and Arild Faxvaag</i>	
Generating Surveillance Data for Nosocomial Infections from Routine Charting in Intensive Care Units	1295
<i>Achim Michel-Backofen, Dorothee Steinke, Christian Katzer, Florian Brenck and Andreas Böning</i>	
Great Help for Small People – The Development of a Children’s Emergency App	1297
<i>Judith Born, Christina Ohlmeier, Jasmina Kis, Ralf Sick, Jörg Oberfeld and Christian Juhra</i>	
Greek Hospitals Web Accesibility	1299
<i>Andriana Magdalinou and John Mantas</i>	
Health Facility Ownership Type and Performance on HIV Indicator Data Reporting in Kenya	1301
<i>Milka B. Gesicho, Ankica Babic and Martin C. Were</i>	

Hierarchical Clustering for Image Classification in Dermatology: Towards Mobile Deploying	1303
<i>Miruna D. Ciulu, Stefan Holban and Diana Lungeanu</i>	
How to Decide upon Nursing Technologies – A Participation-Based Approach	1305
<i>Jörn Krückeberg, Ronny Klawunn, Inessa Fuge, Reza Mazhari, Regina Schmeer and Nicole Hechtel</i>	
How to Extract and Explore Big Data for Fraud Detection in the Healthcare Sector: The EOPYY Case Study	1307
<i>Spiros Georgakopoulos, Parisis Gallos and Vassilis Plagianakos</i>	
Identification of EMCI in MR Brainstem Structure Using Fractal Measures and Random Forest Approach	1309
<i>Rohini Palanisamy and Ramakrishnan Swaminathan</i>	
Identifying and Assessing Competencies for Staff Working in Public Health Emergencies	1311
<i>Heini Utunen, Andrew Black, Oliver Stucke, Melissa Attias and Gaya M. Gamhewage</i>	
Identifying Experts Reasoning in Antibiotic Treatment with Preference Learning	1313
<i>Karima Sedki, Chaymae Lakraflı, Jean-Baptiste Lamy and Rosy Tsopra</i>	
Identifying Indicators to Assess and Monitor Data Integration Engines Systems	1315
<i>Priscila Maranhao, Ana Margarida Pereira and Ricardo J. Cruz-Correia</i>	
Implementation of Wireless Biosensor for Continuous EEG Monitoring in Neurological Intensive Care	1317
<i>Tien-Yu Hsu, Terry B.J. Kuo and Ming-Chin Lin</i>	
Implications of Clinicians' Attitudes Towards Clinical Decision Support Systems	1319
<i>Anshu Ankolekar, Britt van der Heijden, Andre Dekker, Cheryl Roumen, Dirk de Ruysscher, Bart Reymen, Ruud Houben, Peter Fick, Sander Puts, Joeri Veugen, Adriana Berlanga, Cary Oberije and Rianne Fijten</i>	
Incorporating Tacit Knowledge of Experts in the Assessment of Shelters Under Disaster	1321
<i>Masaharu Nakayama and Tadashi Ishii</i>	
Integrating Socially Assistive Robots into Japanese Nursing Care	1323
<i>Markus Kolstad, Natsu Yamaguchi, Ankica Babic and Yoko Nishihara</i>	
Introducing Clinical Informatics Course in Medical School Curricula: Lessons Learned from Medical Faculty University of Belgrade	1325
<i>Jelena Milin-Lazovic, Andja Cirkovic, Marko Savic, Ognjen Milicevic, Ljubica Carevic, Nikola Ilic, Dejana Stanisavljevic and Natasa Milic</i>	
Knowledge Driven Phenotyping	1327
<i>Honghan Wu, Minhong Wang, Qianyi Zeng, Wenjun Chen, Thomas Nind, Emily Jefferson, Marion Bennie, Corri Black, Jeff Z. Pan, Cathie Sudlow and Dave Robertson</i>	

PoleSat's Modelling a Reorganization of Vascular Catheterization Provision at the Territorial Hospital Grouping Level	1353
<i>Anne Quesnel-Barbet, Julien Soula, Erik-André Sauleau, Pierre Parrend, Pierre Bazile, François Dufosse, Gilles Maignant, Pascal Staccini, Frédéric Albert and Arnaud Hansske</i>	
Predicting Diagnosis Code from Medication List of an Electronic Medical Record Using Convolutional Neural Network	1355
<i>Jakir Hossain Bhuiyan Masud and Ming-Chin Lin</i>	
Prediction of ROSC After Cardiac Arrest Using Machine Learning	1357
<i>Nan Liu, Andrew Fu Wah Ho, Pin Pin Pek, Tsung-Chien Lu, Pairoj Khruekarnchana, Kyoung Jun Song, Hideharu Tanaka, Ghulam Yasin Naroo, Han Nee Gan, Zhi Xiong Koh, Huei-Ming Ma and Marcus Ong</i>	
Predictors of Hospital Admissions Based on Routinely Collected Everyday Observations in a Community Care Setting	1359
<i>Mads Nibe Stausholm, Pernille Heyckendorff Secher and Ole Kristian Hejlesen</i>	
Preliminary Evaluation of a mHealth Coaching Conversational Artificial Intelligence for the Self-Care Management of People with Sickle-Cell Disease	1361
<i>David-Z. Issom, Jessica Rochat, Gunnar Hartvigsen and Christian Lovis</i>	
Preliminary Evaluation of Market Mechanism-Based Bed Allocation System	1363
<i>Shusuke Hiragi, Osamu Sugiyama, Jun Hatanaka, Shosuke Ohtera, Goshiro Yamamoto, Kazuya Okamoto, Masayuki Nambu and Tomohiro Kuroda</i>	
Preliminary Results from a Pragmatic Clinical Trial of MyPEEPS Mobile to Improve HIV Prevention Behaviors in Young Men	1365
<i>Rebecca Schnall, Lisa Kuhns, Cynthia Pearson, Josh Bruce, D. Scott Batey, Asa Radix, Uri Belkind, Marco A. Hidalgo, Sabina Hirshfield, Sarah Ganzhorn and Robert Garofalo</i>	
Preliminary Safety Analysis of a Wearable Clinic for the Early Detection of Psychotic Relapse	1367
<i>Ibrahim Habli, Charlotte Stockton-Powdrell, Matthew Machin, Paolo Fraccaro, Shon Lewis and Niels Peek</i>	
Primary Care Prostate Cancer Case Ascertainment	1369
<i>Agnieszka Lemanska, Sara Faithfull, Harshana Liyanage, Sophie Otter, Marina Romanchikova, Julian Sherlock, Nadia A.S. Smith, Spencer A. Thomas and Simon de Lusignan</i>	
Proposal of an Architecture for Terminology Management in a Research Project	1371
<i>Sara Mora, Sumit Madan, Stephan Gebel and Mauro Giacomini</i>	
Providing Personal Health Records Based on a Mobile Application for Emergencies	1373
<i>Su Min Kim, Suyoung Yoo, Won Chul Cha and Tae Rim Kim</i>	

Qualitative Study to Design an Online Community for Patients with Psoriasis <i>Giuliana Colussi, Janine Sommer, Mariana Simón, Lucila Bruchanski and Daniel Luna</i>	1375
Results of the Use of the Teleconsultation Platform After 2 Months of Implementation <i>Emiliano Lopez, Maia Berlin, Romina Stein, Erica Cozzi, Andrea Bermudez, Humberto Mandirola Brieux, Martín Díaz Maffini, Roberto Moldes, Teodorico Bousquet, Patricia MacCulloch, Daniel A. Rizzato Lede, Cintia D. Speranza and Alejandro López Osornio</i>	1377
Risk Factors for Chronic Diabetes Patients <i>Oleg Metzker, Kirill Magoev, Stanislav Yanishevskiy, Alexey Yakovlev and Georgy Kopanitsa</i>	1379
Rosacea Patients Are at Higher Risk for Obstructive Sleep Apnea: Automated Retrospective Research <i>Peter L. Elkin, Sarah Mullin and Sylvester Sakilay</i>	1381
Safety and Drugs: How Do We Record Medication Consumption and Prescription in Electronic Medical Records? A Look on Aspirin <i>Mercedes Volpi, Santiago Esteban and Sergio Terrasa</i>	1383
Service Personalization in IT-Supported Finnish Healthcare <i>Olli Korhonen, Vasiliki Mylonopoulou and Guido Giunti</i>	1385
Serving Health Emergency Responders Through Online Learning – Findings from OpenWHO's Global User Metrics <i>Heini Utunen, Ngouille Ndiaye, Ursula Zhao, Richelle George, Melissa Attias and Gaya Gamhewage</i>	1387
Smartphone Applications for Range of Motion Measurement in Clinical Practice: A Systematic Review <i>Anne-Violette Bruyneel</i>	1389
Social Media Mining for Postpartum Depression Prediction <i>Alina Trifan, Dave Semeraro, Justin Drake, Radek Bukowski and José Luís Oliveira</i>	1391
Specification and Distribution of Vocabularies Among Consortial Partners <i>Mirko Gruhl, Ines Reinecke and Martin Sedlmayr</i>	1393
Standardized Communication Using FHIR and SNOMED CT in Treatment of Diabetic Foot Syndrome Within the Project iFoot <i>Dominik Ludmann, Elisabeth Pantazoglou and Hubert Otten</i>	1395
Strategy for the Analysis and Visualization of Electronic Medical Record Data for Public Hospitals in the City of Buenos Aires <i>Tomas Galluzzi, Marina Ridao and Santiago Esteban</i>	1397
Strengthening Behavior and Social Functioning Among Persons with Autism Spectrum Conditions Using Artificial Intelligence and Behavioral Activation: Protocol for the Well-Being and Health for Loved onEs with ASD (WHOLE) Psychosocial Pilot Randomized Controlled Trial <i>Pierre Bowrin and Usman Iqbal</i>	1399

TEDIS, a Comprehensive Data Model for In-Depth Clinical Assessment of Patients Affected with Neuro-Developmental Disorders Including Autism <i>Mohamed Ben Said, Laurence Robel, Pauline Chaste, Didier Perisse, Marie-Joelle Oreve, Pascale Zylberberg, Anne Philippe, Catherine Jousselman, Stéphanie Lacoste, Mario Speranza, Ines Hafsa, Fatouma Cisse, Zoubair Cherqaoui and Jean-Philippe Jais</i>	1401
Telemedicine Interventions for the Management of Diabetes: A Systematic Review and Meta-Analysis <i>Stine Hangaard, Sisse H. Laursen, Flemming W. Udsen, Peter Vestergaard and Ole Hejlesen</i>	1403
The 'Back Office' of a Dispensing Cabinet: Technology and Work Contributing to Medication Safety <i>Valentina Lichner, Mirela Prgomet, Bryony Dean Franklin and Johanna I. Westbrook</i>	1405
The Prevalence of Alcohol and Tobacco Use Associated Risk Factor of Noncommunicable Diseases in Si Sa Ket Province, Thailand <i>Kampol Khemthong, Niruwan Turnbull, Savittri Rattanopad Suwanlee and Karl Peltzer</i>	1407
Toward a Harmonized WHO Family of International Classifications Content Model <i>Samson W. Tu, Csongor I. Nyulas, Tania Tudorache, Mark A. Musen, Andrea Martinuzzi, Coen van Gool, Vincenzo della Mea, Christopher G. Chute, Lucilla Frattura, Nick Hardiker, Huib ten Napel, Richard Madden, Ann-Helene Almborg, Jeewani Anupama Ginige, Catherine Sykes, Can Cekik and Robert Jakob</i>	1409
Towards an Ontology for Trustful mHealth Apps <i>Hanane Derbouz Rouibete and Stefano Bonacina</i>	1411
Towards Prediction of Type 1 Diabetes Patients Who Fail to Achieve Glycemic Target <i>Morten Hasselstrøm Jensen, Simon Cichosz, Ole Hejlesen, Irl.B. Hirsch and Peter Vestergaard</i>	1413
Translating openEHR Models to FHIR <i>Georg Fette, Max Ertl and Stefan Störk</i>	1415
Uncovering Care Barriers Experienced by Patients and Loved Ones in the Emergency Department: Can a Health App Help? <i>Azzra Mangalji, Kendall Ho, Wayne Choi, Chad Kim Sing, Heather Lindsay, Christopher Wai-Chung Lee, Kent Loftsgard, Michael Lim and Nooshin Jafari</i>	1417
Usability Issues with Mechanical Ventilation Devices <i>Beatrice Coldewey, Annette Diruf, Rainer Röhrig and Myriam Lipprandt</i>	1419
User Satisfaction with a Speech-Enabled Translator in Emergency Settings <i>Antony Akash Janakiram, Johanna Gerlach, Alyssa Vuadens-Lehmann, Pierrette Bouillon and Hervé Spechbach</i>	1421

User-Centered Design for Promoting Patient Engagement in Chronic Diseases Management: The Development of CONCERTO+	1423
<i>Marie-Pierre Gagnon, Mame-Awa Ndiaye, Alain Larouche, Guylaine Chabot, Christian Chabot, Ronald Buyl, Jean-Paul Fortin, Anik Giguère, Annie Leblanc, France Légaré, Aude Motulsky, Claude Sicotte, Holly O. Witteman, Éric Kavanagh, Frédéric Lépinay, Jacynthe Roberge, Hina Hakim, Myriam Brunet-Gauthier, Carole Délétroz, Samira A. Rahimi, Jack Tchuente and Maxime Sasseville</i>	
Using Long Short-Term Memory (LSTM) Neural Networks to Predict Emergency Department Wait Time	1425
<i>Nok Cheng and Alex Kuo</i>	
Subject Index	I
Author Index	XIII

This page intentionally left blank

Section 1

Biomedical Data, Tools and Methods

This page intentionally left blank

A FHIR-Based eConsent App for the Digital Hospital

Manuela WEBER ^{a,b,1}, Alexandra GRIESSBACH ^{c,d}, Regina GROSSMANN ^c and
Jürg BLASER ^{a,b}

^a Faculty of Medicine, University of Zurich, Switzerland

^b Directorate Research and Education, University Hospital Zurich, Switzerland

^c Clinical Trials Center, University Hospital Zurich, Switzerland

^d Faculty of Science, University of Zurich, Switzerland

Abstract. Research projects with humans is a highly regulated field that is currently undergoing rapid changes due to developments in eHealth and mHealth. While a patients data and samples must be thoroughly protected, they are also an invaluable source for fundamental and cutting edge research. There are processes in place to obtain a patient's consent for the use of their data and samples for research. These approaches could be more flexible, user-friendly and modernised. There is a high demand among all parties for a unified, yet differentiated, dynamic and personalised eConsent. An Android app has been developed that brings any existing consent form to mobile devices, including the integration of the process into existing hospital IT using established data standards, such as FHIR and the ResearchStack open source framework. The app is user-tested and shown to work in a hospital setting. Lack of identification and legal drawbacks were determined as the main obstacles for immediate implementation.

Keywords. eConsent, FHIR, Android, ResearchStack, digital hospital, eHealth

1. Introduction

Research projects with humans is a complex and highly regulated field [1], evolved through insight gained by trial and error over the recent decades. This research is subject to new challenges, where large collections of health data and biological samples stored in biobanks and databases, come in conflict with current ethical, legal, and societal implications. The number and size of these repositories is set to increase significantly over the coming years [2].

In Switzerland, the Human Research Act (HRA) [1] allows the use of these data and samples when consent by the patient is present. Consent may be given in form of the traditional *informed consent* (IC), tailored to a specific project, or *general consent* (GC), in which patients allow the use of their data and samples for future, previously unspecified research projects. The HRA as well as the data protection laws give strict instruction on how to handle these highly sensitive medical data and samples.

¹Corresponding Author: Manuela Weber; E-mail: manuela.weber@uzh.ch.

Recent developments in eHealth and mHealth research applications show demand for a consent system, attainable on a patient's own device (eConsent)[3], which may prove profitable to both patient and institution. An eConsent process can simplify a hospital's consent process, as well as allow patients to consent and withdraw at any time; define the terms and conditions for projects using their data and samples (*dynamic consent*); and help patients investigate projects their data and samples are being used in. Simplified hospital processes and increased patient autonomy are not the only benefits of eConsent; Consent forms are often very complex and riddled with medical or legal jargon, which may be better explained through explanatory images and videos. Electronic consent processes can inform patients sufficiently about research studies [4], as interactive elements of eConsent lead to better understanding and recall of key information [5,6]. Additionally, Switzerland is a multi-cultural country with four native languages, this excluding many other languages present due to migration. There is a dire need for quick, accessible and up-to-date consent forms in different languages.

The eConsent concept is not without its challenges. It is well known that older patients visit the hospital more frequently and make out a large sub-population who may have difficulties navigating the application. Furthermore, it may be seen that the eConsent is treated less seriously than a paper-based consent, even though both are legally binding. Nevertheless, current developments in the health system demand for more dynamic, differentiated, and personalised consent solutions.

The prototypes used in the various studies were rather pragmatic and isolated solutions [7] and when authors mentioned the chosen platform, it was mostly iOS based [5,6,8]. This may be due to Apple's initial lead in providing programming tools for the health sector i.e. ResearchKit (<http://researchkit.org>). Android developers, however, caught up by releasing ResearchStack (<http://researchstack.org>). There are now established data standards like FHIR (<https://www.hl7.org/fhir>) and open source frameworks available to create a unified eConsent process for different needs on both platforms, such as C3-PRO (<https://github.com/C3-PRO>).

In this context, the goal of this paper is the research and development of an eConsent technology available on Android and exploring the user acceptance in different patient populations. To our knowledge, we are one of the first team of researchers not only equipped with a prototype, but an eConsent app user tested and readily applicable. The project focus is technological feasibility, user acceptance, and reusability of existing paper consent forms. Our work concentrates on a simple design which allows novice programmers to set up eConsent processes for any paper-based consent form.

2. Methods

Our research consists of two main parts: the design of an easily constructed app and the evaluation of the app's acceptance amongst patients.

In the first step, we conducted a literature search of previous attempts and experiences with eConsent (PubMed, Google Scholar), deciding on the implementation environment. We chose the platform for the app and determined existing frameworks to make the implementation as easy and versatile as possible. The app was then developed, incorporating feedback from different stakeholders, such as thorough research into the technical and legal possibilities and challenges of connecting it to the existing IT environment of the hospital.

For the user acceptance evaluation, the consent forms for the collection of data and samples was integrated into the app and evaluated with patients of the dermatology and the visceral and transplant surgery department. So far, 52 patients were asked to participate of which 33 patients agreed. Patients of different ages were included whether or not they had previously signed the paper-based form.

The administrator selected the document on the eConsent app, scanned the patient label or filled in the patient data manually. The administrator gave a brief standardised introduction and handed the tablet to the patient. The patient read through the document by clicking through the frames, followed by their consent or refusal. When refusing, the process ended and returned to the home screen, while consenting led to a frame where patients signed on screen with a stylus pen. After signing and clicking "next", they returned the tablet to the administrator. In the last step, patients filled out a user acceptance evaluation questionnaire consisting of 9 questions concerning their age group, comprehension, user friendliness, readability, signature, font size, amount of content per frame, preference for eConsent or paper consent as well as general comments and overall experience. Due to current legal restrictions, a paper-based consent form always accompanied the electronic version.

3. Results

3.1. Implementation

With over 80% of mobile devices running worldwide on Android [9] with no restrictive licensing requirements, Android was our operating system of choice. The eConsent app was programmed in Android Studio using Java code. ResearchStack for Android offered an easy to use open source framework for implementation of research apps, including a user interface for a consent process that allows the signing of a consent document on the screen of mobile devices. Integration of the app into existing clinical IT infrastructure was made possible by C3-PRO, an open source toolchain to connect ResearchStack apps to clinical research IT infrastructure through the established standard for medical data FHIR [10].

The consent document, structured into subsections defined and provided by the ResearchStack, is stored in a json file as a FHIR contract resource. Among others, it provides sections for introductions, study information, privacy, usage of collected data and information on withdrawal of consent. Each section is displayed on screen for the patient to read and scroll through. Every section can show a "learn more" link, which navigates the patient to a different screen for more information. The json file containing the consent document is loaded into the app using the C3-PRO framework and passed on to the ResearchStack framework to be presented to the user. To start the consent process, a custom label reader was implemented to read a standard patient bar code and retrieve patient data needed for the consent document, such as name, date of birth and patient ID. The patient is then guided through the sections of the consent document and asked to agree to the terms by signing his signature on the screen of the device. After signing, a PDF of the consent form, complete with the patient information and signature is created and transmitted to the hospital IT system by the C3-PRO framework, either as a FHIR resource or as plain PDF.

3.2. User Acceptance Evaluation

The designed app can be set up for different consent forms with minimal effort. It can run on different Android devices and adjusts to different screen sizes. For the user acceptance study, the app was installed on a 10.5” Samsung tablet. From all 33 patients who completed the consent process and questionnaire, the app was generally positively received and found easy to navigate. The eConsent was generally determined superior over the paper-based version. Results of the questionnaire can be seen in Table 1. A few issues arose during the eConsent process. In some cases, the navigation buttons for the app were not big enough, the amount of text per section was too much and some patients expressed the need for a bigger signature area. Some users also experienced difficulties when resting the palm on the screen when signing. Over all, patients reflected positively about their experience with a few minor and adaptable issues.

Table 1. User Acceptance Questionnaire Results (n=33)

characteristic	comprehension	user friendliness	readability	signature
mean score*	4.6	4.2	4.2	3.2
improvement	integration of media, adjust text per frame	increase button size, swipe function	zoom function, larger font size	increase signature platform

*1 lowest score, 5 highest score

4. Discussion

This project creates a solution that allows a standardised eConsent process for various types of consents with minimal effort or programming knowledge. Due to the use of established data standards, the proposed app can easily be integrated into existing hospital IT infrastructures. The user acceptance evaluation has shown that patients of all ages navigated the technology easily and even prefer the digital version. If further requirements arise in the future, the technology offers plenty of options for adaptation and improvement.

Possible integration of different media types could allow for a more differentiated, dynamic and personalised experience for patients with different preferences. Numerous field tests show that there is wide acceptance and preference of eConsent among patients, medical professionals and researchers [4], which is corroborated by our results. Electronic forms instead of paper forms can lead to greater satisfaction for all parties [7] and field tests have shown that eConsent result in the same or higher consent rate [11,12].

Although the app is ready for implementation, there are some challenges concerning the integration of the app into the IT infrastructure of the institution. The main concern being legal and safety matters. Swiss law does not accept electronic hand-written signatures, with exception of the postal service. eSignatures such as Swiss Post’s SwissID can be costly, are not widely accepted, and do not easily integrate with third party software, such as ResearchStack. No Federal law on recognized eidentification is yet in force. Furthermore, data protection laws are very strict about patient data being stored on mobile devices and transmitted through potentially vulnerable networks. However, despite strict laws and best efforts, eHealth data has become more attractive to hackers over the recent years and are stored in less secure systems [13,14,15]. Substantial initial investment

would be necessary to establish a secure yet accessible infrastructure for a secure unified eConsent system.

5. Conclusion

It is time to supplement or replace paper-based consent forms with a unified yet dynamic and personalised eConsent process. The solution presented is accepted among patients and researchers, and easily integrated into existing and future health networks. The lack of options for eIdentification and legal drawbacks were determined as the main obstacle for immediate implementation.

References

- [1] Die Bundesversammlung der Schweizerischen Eidgenossenschaft. Bundesgesetz über die Forschung am Menschen [Internet]; 2011. Cited 2019 Sept 11. Available from: <https://www.admin.ch/opc/de/classified-compilation/20061313/index.html>.
- [2] Astrin JJ, Betsou F. Trends in biobanking: a bibliometric overview. *Biopreserv Biobank*. 2016 Feb;14(1):65–74. doi:10.1089/bio.2015.0019.
- [3] Bergmann J, Bott O, Pretschner D, Haux R. An e-consent-based shared EHR system architecture for integrated healthcare networks. *Int J Med Inform*. 2007 Feb;76(2-3):130–6. doi:10.1016/j.ijmedinf.2006.07.013.
- [4] Grady C, Cummings SR, Rowbotham MC, McConnell MV, Ashley EA, Kang G. Informed consent. *N Engl J Med*. 2017 Mar;.
- [5] Rowbotham MC, Astin J, Greene K, Cummings SR. Interactive informed consent: randomized comparison with paper consents. *PLoS One*. 2013 Mar;8(3):1–6. doi:10.1371/journal.pone.0058603.
- [6] Rothwell E, Wong B, Rose NC, Anderson R, Fedor B, Stark LA, et al. A randomized controlled trial of an electronic informed consent process. *J Empir Res Hum Res Ethics*. 2014 Dec;9(5):1–7. doi:10.1177/1556264614552627.
- [7] Warriner AH, Foster PJ, Mudano A, Wright NC, Melton ME, Sattui SE, et al. A pragmatic randomized trial comparing tablet computer informed consent to traditional paper-based methods for an osteoporosis study. *Contemp Clin Trials Commun*. 2016 Aug;3:32–8. doi:10.1016/j.conctc.2016.02.003.
- [8] Vanaken H. eConsent study provides insights to shape industry adoption. *Applied Clinical Trials* [Internet]. 2016 08;25(8). [cited 2019 Sept 21]. Available from: <http://www.appliedclinicaltrialsonline.com/econsent-study-provides-insights-shape-industry-adoption>.
- [9] International Data Corporation. Smartphone Market Share; 2019. Cited 2019 Sept 11. Available from: <https://www.idc.com/promo/smartphone-market-share/os>.
- [10] Pfiffner PB, Pinyol I, Natter MD, Mandl KD. C3-PRO: connecting ResearchKit to the health system using i2b2 and FHIR. *PLoS One*. 2016 Mar;11(3):1–8. doi:10.1371/journal.pone.0152722.
- [11] Maurer J, Saccilotto R, Pauli-Magnus C. E-general consent: development and implementation of a nationwide harmonised interactive electronic general consent. *Swiss Medical Informatics*. 2018 Sep;doi:10.4414/smi.34.00412.
- [12] Boutin N, Mathieu K, Hoffnagle A, Allen N, Castro V, Morash M, et al. Implementation of electronic consent at a biobank: an opportunity for precision medicine research. *J Pers Med*. 2016 Jun;6(2):17. doi:10.3390/jpm6020017.
- [13] The University of Illinois at Chicago. Protecting Patient Information in the Age of Breaches; 2016. Cited 2019 Sept 22. Available from: <https://healthinformatics.uic.edu/blog/protecting-patient-information>.
- [14] Liu V, Musen MA, Chou T. Data breaches of protected health information in the United States. *JAMA*. 2015 Apr;313(14):1471–3. doi:10.1001/jama.2015.2252.
- [15] Wikina SB. What caused the breach? An examination of use of information technology and health data breaches. *Perspect Health Inf Manag*. 2014 Oct;11(Fall):1–16.

A Secure Multi-Party Computation Protocol for Time-To-Event Analyses

Lennart VOGELSANG ^{a,b,1}, Moritz LEHNE ^c, Phillipp SCHOPPMANN ^{a,b}, Fabian PRASSER ^{c,d}, Sylvia THUN ^{c,d,e}, Björn SCHEUERMANN ^{a,b} and Josef SCHEPERS ^c

^a Department of Computer Science, Humboldt-Universität zu Berlin, Germany

^b Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin, Germany

^c Berlin Institute of Health (BIH), Germany

^d Charité – Universitätsmedizin Berlin, Germany

^e Hochschule Niederrhein – University of Applied Sciences, Krefeld, Germany

Abstract. The cryptographic method *Secure Multi-Party Computation* (SMPC) could facilitate data sharing between health institutions by making it possible to perform analyses on a “virtual data pool”, providing an integrated view of data that is actually distributed – without any of the participants having to disclose their private data. One drawback of SMPC is that specific cryptographic protocols have to be developed for every type of analysis that is to be performed. Moreover, these protocols have to be optimized to provide acceptable execution times. As a first step towards a library of efficient implementations of common methods in health data sciences, we present a novel protocol for efficient time-to-event analysis. Our implementation utilizes a common technique called *garbled circuits* and was implemented using a widespread SMPC programming framework. We further describe optimizations that we have developed to reduce the execution times of our protocol. We experimentally evaluated our solution by computing Kaplan-Meier estimators over a vertically distributed dataset while measuring performance. By comparing the SMPC results with a conventional analysis on pooled data, we show that our approach is practical and scalable.

Keywords. secure multi-party computation; healthcare data; patient privacy; operational confidentiality of information

1. Introduction

The increasing abundance of digital health data holds great potential for medicine and healthcare. Paired with standardized data formats and analytic tools, this wealth of data can help to improve diagnosis, treatment and prevention as well as efficiency of care. However, health data is distributed across a large number of institutions, and the knowledge to be gained from these isolated data repositories is limited. This has led to the creation of a wide range of national and international data sharing initiatives. But laws and regulations such as the EU General Data Protection Regulation often prevent the disclosure and sharing of health data for privacy reasons.

Secure Multi-Party Computation (SMPC) is a cryptographic approach that can help to overcome this limitation and facilitate data sharing between health institutions. SMPC

¹ Corresponding Author, Lennart Vogelsang, Department of Computer Science, Humboldt-Universität zu Berlin, Rudower Chaussee 25, 12489 Berlin, Germany; E-mail: vogelsal@informatik.hu-berlin.de.

makes it possible to perform analyses on a *virtual data pool*, providing an integrated view of data that is actually distributed – without any of the participants having to disclose their private data [1].

Prior work on SMPC in healthcare includes a protocol for genome-wide association studies that keeps underlying geno- and phenotypes private [2], as well as a study that uses SMPC to investigate drug-target interactions in large pharmacological datasets [3]. Other applications include a disease surveillance system that ensures patient and provider privacy [4] and a method for privacy-preserving health record linkage [5].

2. Objective

Despite its potential as a privacy-preserving approach to health data sharing, SMPC is not yet widely used in health analytics. One reason for this is that SMPC is computationally demanding, and protocols therefore have to be highly optimized for acceptable performance.

As a first step towards a library of efficient SMPC implementations of common methods in health data sciences, we studied the problem of time-to-event analysis over vertically distributed data. In this scenario, one participant (e.g. a hospital) holds relevant clinical data (e.g. the primary diagnosis) while another participant (e.g. a medical practice or death register) holds data on relevant events (e.g. time of death or (re-)admission to hospital) of the same patient group. As a showcase, we present an efficient SMPC implementation of the Kaplan-Meier estimator for survival probabilities.

3. Methods

3.1. Secure Multi-Party Computation

Our implementation uses the Obliv-C framework [6], which provides an abstraction layer for the *garbled circuits* technique. With this widespread approach, two-parties can evaluate a common function without disclosing their inputs by mapping the function to a Boolean circuit [1]. Furthermore, we used a library providing additional primitives for sorting and secure memory that hides access patterns (ORAM) [7–9].

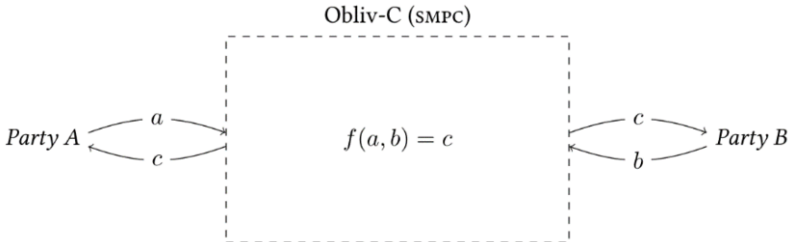


Figure 1. Example for the abstraction layer provided by Obliv-C.

Obliv-C provides an abstraction layer that imitates a Trusted Third Party (TTP). As illustrated in Figure 1, both parties A and B push their data a and b to the TTP, which calculates the function f and returns the output c . In reality, however, there is no TTP;

the computation and its inputs are encrypted and computations are executed distributively between both parties.

3.2. Efficient SMPC Protocol for Kaplan-Meier Estimators

One major drawback of SMPC is its computational overhead. To reduce this overhead, we split our problem into two phases: In the first, SMPC-based phase, computation is minimized and as much computation as possible is offloaded to the second, non-SMPC-based phase. As intermediate results from the SMPC program are revealed to the subsequent non-SMPC programs, special care has to be taken to make sure that these intermediate results do not leak private information.

In the particular case of the Kaplan-Meier estimator, the survival probabilities of individual patients should not be leaked, as this would create a direct mapping between the result and each patient's private information. Instead, values have to be aggregated so that only cumulated distinct data points are revealed. To aggregate time points in the SMPC phase, we used the secure memory (ORAM) mentioned above.

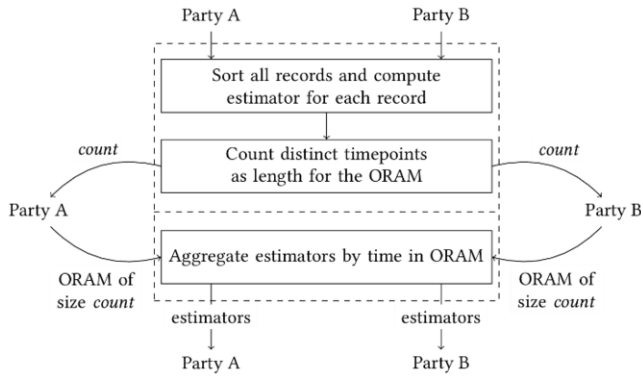


Figure 2. Schematic view of our algorithm implementing the Kaplan-Meier estimator.

One challenge with this approach is that the ORAM has to be initialized with a maximum size. While the number of patient records could be used as an upper bound, this adversely impacts performance, as ORAMs do not scale well. However, the length of the result is typically lower because time points get aggregated and not every patient has an event. The size of the ORAM can thus be limited to the number of distinct time points. We achieve this by iterating over the dataset in two (sub-) phases: first, we aggregate the values without storing them, only counting the number of distinct time points, revealing the count afterwards; then we initialize the ORAM and perform the calculations again, this time storing the results as shown in Figure 2. This does not impact privacy in any way, as the count is leaked at the end of the protocol anyway.

The source code of our SMPC protocol for Kaplan-Meier estimators is available at <https://github.com/LevitatingOrange/secure-kaplan-meier>.

3.3. Experimental Setup

In our experiments we used two datasets: (1) A synthetic dataset including patient demographics (date of birth, gender and postal code), case information (case number,

admission and discharge dates), coded diagnoses (ICD-10-GM) and procedures (German OPS codes); (2) a simulated dataset from a death registry. The datasets included a total of 3,927 patients with 5,110 cases. The patients shared a common identifier across both datasets (e.g. a social security or health insurance number).

We performed two experiments: (1) a “conventional” analysis in R (www.R-project.org) on the pooled datasets on one machine; (2) an analysis on distributed data with our SMPC protocol simulating two machines connected with a local area network (LAN, bandwidth: 119 Mbits/s, average latency: 0.347ms) and a wide area network (WAN, bandwidth: 31.3 Mbits/s, average latency: 5.11ms). We used two laptops for the LAN scenario: a six core Intel Core i7-8850H processor and 32 GB RAM machine and another with a four core Intel Core i7-4960HQ and 16 GB RAM. In the WAN scenario, we connected the former laptop to a virtual machine on a hosting provider with 12 GB RAM and a six core Intel Xeon E5-2680 processor.

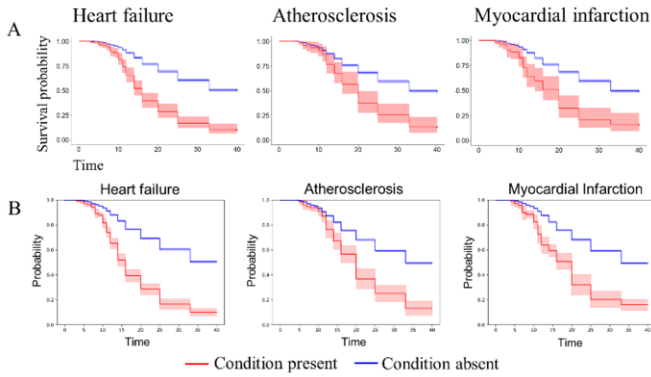


Figure 3. Kaplan-Meier plots obtained using (A) pooled data and (B) the SMPC protocol.

4. Results

Figure 3 shows the Kaplan-Meier graphs comparing patients with (vs. without) a diagnosis of heart failure (ICD-10 code: I50*), atherosclerosis (ICD-10 code: I70*) and myocardial infarction (ICD-10 code: I21*). As can be seen, the results obtained in both experiments were in high agreement.

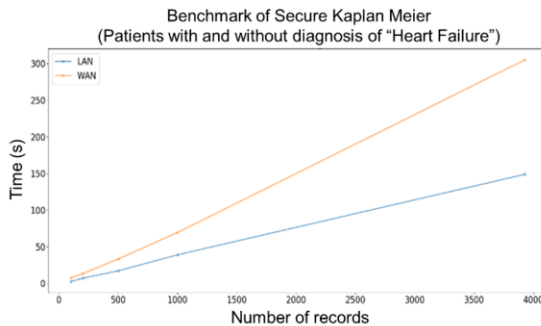


Figure 4. Execution times of the SMPC Kaplan-Meier estimator in the LAN and WAN scenario.

Figure 4 shows the execution times of the SMPC protocol. Both scenarios (LAN and WAN) showed a linear increase in computation time in relation to the number of records processed. The complete dataset of nearly 4,000 records was processed in about five minutes in the WAN scenario.

5. Discussion and Conclusion

In this article, we have presented an efficient SMPC implementation of the Kaplan-Meier estimator over vertically distributed data.

Slight differences in the survival probabilities resulting from the SMPC and the conventional implementation can be explained by the fact that Obliv-C only supports 32-bit single precision floating point arithmetic (rather than the 64-bit precision floating-point operations of the conventional analysis). Moreover, the experiments showed that our approach provides practical execution times. Even with very large datasets (>1,000,000 records), we expect computations to take not more than several hours or maximally a few days. This is acceptable in typical research projects, especially when considering that the alternative, in many cases, would be not to do the project at all.

In future work, we plan to implement additional SMPC protocols for common statistical methods in health analytics and test them in real-world scenarios. An interesting candidate would be an SMPC implementation of a Cox proportional hazard model for more advanced statistical analyses of survival times. In the long-term, SMPC protocols could be made available in an open source library, providing a collection of methods that researchers can choose from for their specific projects.

For long, SMPC remained a largely academic field, especially due to its sub-par performance compared with traditional analysis methods. With recent progress towards efficient SMPC protocols and the focus on data privacy and security, SMPC can play a large role in advancing analysis across distributed healthcare data.

References

- [1] A.C. Yao, How to generate and exchange secrets, in: 27th Annu. Symp. Found. Comput. Sci. SfcS 1986, 1986: pp. 162–167.
- [2] H. Cho, D.J. Wu, and B. Berger, Secure genome-wide association analysis using multiparty computation, *Nat. Biotechnol.* **36** (2018) 547–551.
- [3] B. Hie, H. Cho, and B. Berger, Realizing private and practical pharmacological collaboration, *Science*. **362** (2018) 347–350.
- [4] K. El Emam, J. Hu, J. Mercer, L. Peyton, M. Kantarcioglu, B. Malin, D. Buckeridge, S. Samet, and C. Earle, A secure protocol for protecting the identity of providers when disclosing data for disease surveillance, *J. Am. Med. Inform. Assoc. JAMIA*. **18** (2011) 212–217.
- [5] F. Chen, X. Jiang, S. Wang, L.M. Schilling, D. Meeker, T. Ong, M.E. Matheny, J.N. Doctor, L. Ohno-Machado, and J. Vaidya, Perfectly Secure and Efficient Two-Party Electronic-Health-Record Linkage, *IEEE Internet Comput.* **22** (2018) 32–41.
- [6] S. Zahur, and D. Evans, Obliv-C: A Language for Extensible Data-Oblivious Computation, *ACR Cryptol. EPrint Arch.* (2015) 20.
- [7] O. Goldreich, and R. Ostrovsky, Software Protection and Simulation on Oblivious RAMs, *J ACM*. **43** (1996) 431–473.
- [8] S. Zahur, X. Wang, M. Raykova, A. Gascon, J. Doerner, D. Evans, and J. Katz, Revisiting Square-Root ORAM: Efficient Random Access in Multi-party Computation, in: 2016 IEEE Symp. Secur. Priv. SP, IEEE, 2016: pp. 218–234.
- [9] J. Doerner, and A. Shelat, Scaling ORAM for Secure Computation, in: Proc. 2017 ACM SIGSAC Conf. Comput. Commun. Secur. - CCS 17, ACM Press, 2017: pp. 523–535.

A Semantic Similarity Evaluation for Healthcare Ontologies Matching to HL7 FHIR Resources

Athanasios KIOURTIS^{a,1}, Argyro MAVROGIORGOU^a and Dimosthenis KYRIAZIS^a

^a*Department of Digital Systems, University of Piraeus, Greece*

{kiourtis,margy,dimos}@unipi.gr

Abstract. Healthcare 4.0 demands healthcare data to be shaped into a common standardized and interoperable format for achieving more efficient data exchange. Most of the techniques addressing this domain are dealing only with specific cases of data transformation through the translation of healthcare data into ontologies, which usually result in clinical misinterpretations. Currently, ontology alignment techniques are used to match different ontologies based on specific string and semantic similarity metrics, where very little systematic analysis has been performed on which semantic similarity techniques behave better. For that reason, in this paper we are investigating on finding the most efficient semantic similarity technique, based on an existing approach that can transform any healthcare dataset into HL7 FHIR, through the translation of the latter into ontologies, and their matching based on syntactic and semantic similarities.

Keywords. Healthcare, Ontology alignment, Semantic similarity, HL7 FHIR

1. Introduction

In the last decade, there has been a transition from a data-poor to a data-rich world, with the aim of improving the quality of transport, governance, environment, communication and health. Much of this unprecedented increase in data generation can be attributed to the abundance of thousands of mobile devices, wearables, and sensors. Most of these devices are typically characterized by a high degree of heterogeneity, in terms of having different characteristics, capabilities, or network specifications, needing to be easily manageable in particular with regard to the management of the heterogeneous data they generate. Especially for the Internet of Medical Things (IoMT) devices it is undeniable that vast amounts of heterogeneous medical data are becoming available, thus completely reshaping the Healthcare 4.0. In order for these heterogeneous medical data to be exchanged with multiple stakeholders, and to be a key driver of providing personalized and efficient medical care to patients and citizens, interoperability is the only way.

Taking into consideration these challenges, in [11] a holistic approach has been presented for achieving interoperability through the transformation of healthcare data into its corresponding HL7 FHIR structure. Shortly, the provided mechanism was building the healthcare ontologies that were stored into a triplestore, in order to identify

¹ Corresponding Author, Athanasios Kiourtis, 80, M. Karaoli & A. Dimitriou St., 18534 Piraeus, Greece
E-mail: kiourtis@unipi.gr.

and compare their syntactic and semantic similarity with the HL7 FHIR Resources. According to the aggregation of the syntactic and semantic similarity results, the alignment and translation to the HL7 FHIR was taking place. In this paper, we are going to fill the gap of the semantic similarity of the aforementioned mechanism, as there has been little systematic analysis on which semantic similarity techniques perform well when applied to ontology alignment.

The remainder of this paper is organized as follows. In Section 2, some semantic similarity techniques are illustrated, while Section 3 depicts the overall approach. Section 4 includes the evaluation of the derived results, presenting our concluding remarks.

2. Semantic Similarity Techniques

Several methods for determining semantic similarity between terms have been proposed in the literature. Word2Vec [2] is a two-level neural network that processes text. Its input is a body of text and its output is a set of vectors. The purpose and utility of Word2Vec is to group similar word vectors together in the vector space, detecting mathematical similarities. Word2Vec creates vectors that are distributed arithmetic word representations, such as the single word context, without human intervention. Given enough data, usage, and the current environment, Word2Vec can make very accurate guesses about the meaning of a word based on impressions of the current word in other situations in the past. WordNet [3] is a great English based dictionary. Nouns, verbs, adjectives, and adverbs are grouped into sets of synsets, each of which expresses a separate meaning. Synsets are interrelated through conceptual-semantic and lexical relationships. The resulting network of related words and concepts can be navigated with the browser. The structure of WordNet makes it a useful tool for computational linguistics and NLP. WordNet looks superficially like a treasure trove because it groups words according to their meanings. Cortical.io [4] proposes a new way of understanding the natural language that transcends the limitations of other artificial intelligence approaches. Semantic Folding [5] is inspired by the most recent findings on how the brain processes information. This theory introduces the Semantic Fingerprint, which explicitly codifies the concept, including all concepts and frameworks. The system understands the affinity of two elements by measuring the overlap of their fingerprints.

3. Proposed Approach

In the context of ontology alignment of healthcare data, several techniques are presented that lead to better decision making, and interoperable exchange and use of data, providing solutions that are adapted to deliver results under specific scenarios. To address this gap, in [1] a holistic approach is proposed that is capable of identifying common links between healthcare ontologies and the ontologies that represent the HL7 FHIR Resources. The preliminary steps of the developed approach were dealing with the automatic transformation of the healthcare dataset and the HL7 FHIR Resources into their ontological structure. As soon as the ontologies had been constructed, they were stored into a triplestore. Afterwards, two sub-mechanisms had occurred, providing the syntactic and the semantic similarity between two different ontologies, based on their syntactic representations and their semantic knowledge accordingly. The final stage of the mechanism implemented an additional sub-mechanism that aggregated and merged the

outcomes of the aforementioned sub-mechanisms, providing the overall alignment results, for the final transformation of the healthcare dataset into HL7 FHIR. Henceforth based on this developed mechanism, in our case, we assume that healthcare ontologies have already been constructed for both the ingested healthcare dataset and the different HL7 FHIR Resources. Consequently, the three described semantic similarity techniques are going to be implemented, for measuring the semantic similarity of the built ontologies.

3.1. Semantic Similarity Identifier

The objective of the Semantic Similarity Identifier is to provide the means for aligning and matching the different healthcare dataset and HL7 FHIR Resources ontologies, according to their semantic meaning (Fig. 1).

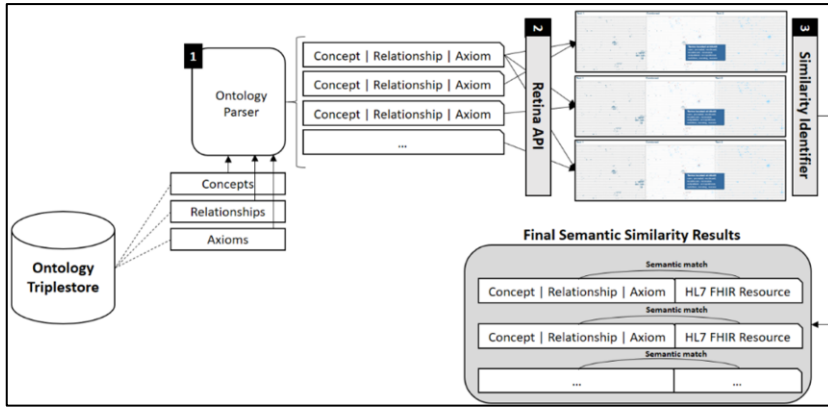


Figure 1. Semantic Similarity Identifier.

4. Evaluation

In our case, in order to evaluate the three different semantic similarity techniques, we are going to implement them under the same use case scenario where we will calculate their precision, recall, F-measure, and total errors [6], with respect to a UMLS-based reference alignment [7]. After this calculation, a comparison between the results will take place in order to conclude to the semantic similarity technique that is more suitable for aligning ontologies and calculating their semantic similarity in the healthcare domain.

4.1. Application of Semantic Similarity Techniques

The use case exploits a healthcare dataset structured in CSV format, covering all the previous steps, consisting of 5000 different instances of the personal information of certain citizens about their: (i) personal identifier (*subject*), (ii) gender (*gender*), (iii) date of birth (*dateOfBirth*), and (iv) cause of death (*causeOfDeath*). After performing the steps introduced into the Semantic Similarity Identifier, the following tables are being created (Table 1, Table 2, and Table 3), depicting the HL7 FHIR Resources with higher semantic similarity degrees, compared with the concept names of the use case dataset, through the three different semantic similarity techniques.

Table 1. Semantic Similarity Identifier Top Results for Cortical.io

Use Case Dataset Attribute	HL7 FHIR Resources	Cortical.io
subject	Patient.identifier	0.947 (~95%)
gender	Patient.gender	0.863 (~86%)
dateOfBirth	Patient.birthDate	0.971 (~97%)
causeOfDeath	Patient.deceased	0.458 (~46%)

Table 2. Semantic Similarity Identifier Top Results for Word2Vec

Use Case Dataset Attribute	HL7 FHIR Resources	Word2Vec
subject	Patient.contact.address	0.357 (~36%)
gender	Patient.gender	0.663 (~66%)
dateOfBirth	Patient.birthDate	0.891 (~89%)
causeOfDeath	-	0 (0%)

Table 3. Semantic Similarity Identifier Top Results for Wordnet

Use Case Dataset Attribute	HL7 FHIR Resources	Wordnet
subject	Patient.gender	0.772 (~77%)
gender	Patient.gender	0.861 (~86%)
dateOfBirth	Patient.multipleBirth	0.561 (~56%)
causeOfDeath	Patient. deceased	0.619 (~62%)

4.2. Conclusions & Discussion of Results

In order to evaluate these results with regards to the metrics of precision, recall, F-measure, and errors, the specific use case dataset was translated manually in HL7 FHIR, so as to compare the results of the developed mechanism, with the actual outcomes. In Fig. 2, a visualization of the aforementioned results is depicted through a bar chart, showing how the three different semantic similarity techniques behave in the total semantic similarity identification, with regards to the manually provided results which did not contain any errors (0% of errors), while their precision, recall and F-measure had the value of 100%.

It is clear that there was not any semantic similarity technique that provided 100% accurate results. As it can be seen through Fig. 2, among the different techniques, the Cortical.io behave better since it provided less errors (13%) compared with the other semantic similarity techniques (31%, and 23% accordingly). Moreover, the F-measure (a result of both the precision and recall) of the Cortical.io had the greatest value (84%) in comparison with the other semantic similarity techniques (77%, and 65% accordingly). Consequently, this lead us to the conclusion that for the Semantic Similarity Identifier the most convenient semantic similarity technique to use is the Cortical.io, since it provides more efficient and reliable results.

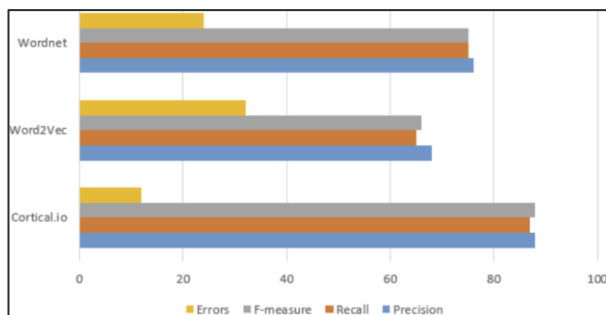


Figure 2. Visualization of Comparison Results.

In the field of healthcare interoperability, semantic similarity has to be identified, for which there has been little systematic analysis on which techniques perform well when applied to ontology alignment. For addressing this gap, in this paper a previous research was considered, about a healthcare ontology matching mechanism that has the ability of transforming healthcare data to HL7 FHIR format. Therefore, based upon this research, in order to identify the required semantic similarity, three of the most well-known semantic similarity techniques were used under the same scenario, in order to result to the most efficient technique of semantic similarity in the cases of healthcare ontologies alignment, and consequently to the most efficient HL7 FHIR transformation.

To this context, we will continue working on the evaluation of the Semantic Similarity Identifier mechanism with additional semantic similarity techniques, to identify the technique that is most suitable for the case of healthcare ontologies alignment. In addition, we will continue on this evaluation with datasets of different sizes, standards and formats, respecting privacy issues, based on the mechanism developed in [8].

Acknowledgment

The research leading to this result has received funding from the European Union's Horizon 2020 research and innovation programmes under grant agreement No 727560 (CrowdHEALTH project) and grant agreement No 826106 (InteropEHRate project).

References

- [1] A. Kiourtis, A. Mavrogiorgou, D. Kyriazis, S. Nifakos, Aggregating the Syntactic and Semantic Similarity of Healthcare Data towards their Transformation to HL7 FHIR through Ontology Matching, *Int. Journal of Medical Informatics*, 132 (2019).
- [2] Y. Goldberg, O. Levy, word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722* (2014).
- [3] Wordnet, <https://wordnet.princeton.edu/>
- [4] Cortical.io, <https://www.cortical.io/>
- [5] F. D. S. Webber, Semantic folding theory and its application in semantic fingerprinting. *arXiv preprint arXiv:1511.08855* (2015).
- [6] A. Kiourtis, A. Mavrogiorgou, A. Menychtas, I. Maglogiannis, D. Kyriazis, Structurally Mapping Healthcare Data to HL7 FHIR through Ontology Alignment. *Journal of medical systems*, 43(3) (2019).
- [7] Ontology Alignment Evaluation Initiative, <http://oaei.ontologymatching.org/2018/>
- [8] A. Kiourtis, A. Mavrogiorgou, D. Kyriazis, Towards a Secure Semantic Knowledge of Healthcare Data Through Structural Ontological Transformations. *Joint Conference on Knowledge-Based Software Engineering*, (2018) 178-188.

A Semi-Automated Approach for Multilingual Terminology Matching: Mapping the French Version of the ICD-10 to the ICD-10 CM

Emmanuelle SYLVESTRE^{abc1}, Guillaume BOUZILLÉ^{ab}, Michael McDUFFIE^d,

Emmanuel CHAZARD^c Paul AVILLACH^d, Marc CUGGIA^{ab}

^aINSERM, LTSI UMR 1099, F-35000, Rennes France

^bCHU Rennes, Centre de Données Cliniques, F-35000, Rennes France

^cCHU Martinique, Centre de Données Cliniques, F-97200, Martinique France

^dDepartment of Biomedical Informatics, Harvard Medical School, Boston, MA 02115 USA.

^eUniversité de Lille, CHU Lille, CERIMEA2694, F-59000 Lille, France.

Abstract. The aim of this study was to develop a simple method to map the French International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10) with the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10 CM). We sought to map these terminologies forward (ICD-10 to ICD-10 CM) and backward (ICD-10 CM to ICD-10) and to assess the accuracy of these two mappings. We used several terminology resources such as the Unified Medical Language System (UMLS) Metathesaurus, Bioportal, the latest version available of the French ICD-10 and several official mapping files between different versions of the ICD-10. We first retrieved existing partial mapping between the ICD-10 and the ICD-10 CM. Then, we automatically matched the ICD-10 with the ICD-10-CM, using our different reference mapping files. Finally, we used manual review and natural language processing (NLP) to match labels between the two terminologies. We assessed the accuracy of both methods with a manual review of a random dataset from the results files. The overall matching was between 94.2 and 100%. The backward mapping was better than the forward one, especially regarding exact matches. In both cases, the NLP step was highly accurate. When there are no available experts from the ontology or NLP fields for multi-lingual ontology matching, this simple approach enables secondary reuse of Electronic Health Records (EHR) and billing data for research purposes in an international context.

Keywords. ICD-10, Clinical terminologies, Interoperability, Multilingual matching

1. Introduction

The International Statistical Classification of Diseases and Related Health Problems, 10th revision (WHO-ICD-10) one of the most popular terminologies used in around the world.

¹ Corresponding Author, *Faculté de médecine, Université Rennes 1, 2 Avenue du Professeur Léon Bernard 35043 Rennes Cedex 9, France*; E-mail: emmasyl@gmail.com.

It is a standard diagnostic terminology created and maintained by the World Health Organization (WHO) since 1990 for diagnostic coding[1]. The French healthcare system uses a French version of the WHO-ICD-10 (Classification Internationale des Maladies, 10e version, CIM-10) since 1997[2], while the United States created and implanted their own adaptation of the terminology (International Classification of Diseases, 10th Revision, Clinical Modification, ICD-10-CM) on October, 1st, 2015[3]. There are far more codes in the ICD-10 CM than in the CIM-10, even though they share the same common denominator: the WHO-ICD-10[3].

Multilingual ontology matching is the process of finding correspondences between ontologies of different languages to allow them to interoperate[4]. This enables secondary reuse of Electronic Health Records (EHR) and billing data from different healthcare systems for research purposes, especially if data from the United States is involved in the study. We can use two main strategies for multilingual ontology matching: direct and indirect alignment. The direct alignment is translation-based and uses external resources to help with translation, while the indirect alignment uses intermediary mappings between the source and target ontologies. Furthermore, mapping two ontologies can be an automated or manual process. Manual mapping is still the prevalent choice for ontology matching, but necessitates a large team of experts, is time-consuming, and is prone to errors[5]. On the other hand, automated approaches use public terminology resources such as the Unified Medical Language System (UMLS)[4] or Biportal[6] but those sources are extremely incomplete outside of the English speaking world[7]. Therefore, when the purpose of a study is not the mapping itself but a necessary step to join databases, it should be possible to overcome the semantic interoperability issue by combining different automated matching techniques to conduct the study, even with limited resources or experts from the ontology matching field.

The aim of this study was to develop and evaluate a simple method to link the French ICD-10 (or any version of the WHO ICD-10) with the ICD-10 CM. We sought to map these terminologies forward (ICD-10 to ICD-10 CM) and backward (ICD-10 CM to ICD-10).

2. Methods

2.1. Terminology resources

Since there were no direct mapping files for our study, we used all the intermediate mapping files available online. We used four data sources: i) the UMLS Metathesaurus, which integrates and assigns a unique identifier to synonymous concepts from several standard biomedical technologies (including ICD-10 CM, WHO-ICD-10 and a 1998 version of the CIM-10)[4], ii) Biportal, which is a comprehensive repository of standard terminologies created by the National Center for Biomedical Ontology (NCBO), with an ontology alignment tool, called Lexical OWL Ontology Matcher (LOOM)[6], iii) the latest version of the CIM-10, which is mainly a translation of the WHO-ICD-10 with more children, and is publicly available on the national billing agency website (Agence Technique de l'Information sur l'Hospitalisation, ATIH)[2] and iv) different existing mapping files with forward and backward mapping such as: the General Equivalent Mapping (GEM) Files from the National Center for Health Statistics (NCHS)[8], the New-Zealand mapping files from the Ministry of Health (ICD-10 Australian Modification and ICD-9 Australian Modification)[9] and the mapping files between the

ICD-10 AM and ICD-10, from the Australian Consortium for Classification Development (ACCD)[10].

2.2. Mapping method

We decided to map only the first 4-character codes of the CIM-10, because the WHO-ICD-10 is mainly 4-character codes, except for chapters XIII, XIX, XX, which means that after this position, there was a bigger risk of mapping codes with a different signification between the two languages.

Our strategy used three main steps (Figure 1). First, we used the NCBO API to retrieve the mapping between the ICD-10 CM and the WHO-ICD-10 and the partial mapping between the CIM-10 and the WHO-ICD-10 from the UMLS Metathesaurus and LOOM algorithm, which is available through the NCBO API. Then, we matched each terminology using the different mapping files mentioned above. And finally, we used manual review and natural language processing (NLP) to recognize labels between the two terminologies with custom R scripts. The NLP process was only used on unmatched codes after the two first steps. For each set, one native-speaking investigator from each language extracted the two or three main words of the ICD or CIM label, then they built together a translation dictionary and used rules-based NLP to match the labels and codes. For the remaining unmatched codes, we mapped them to their three-character codes parent. After all these steps, we considered the code unmatched if we could not find an exact or approximate match.

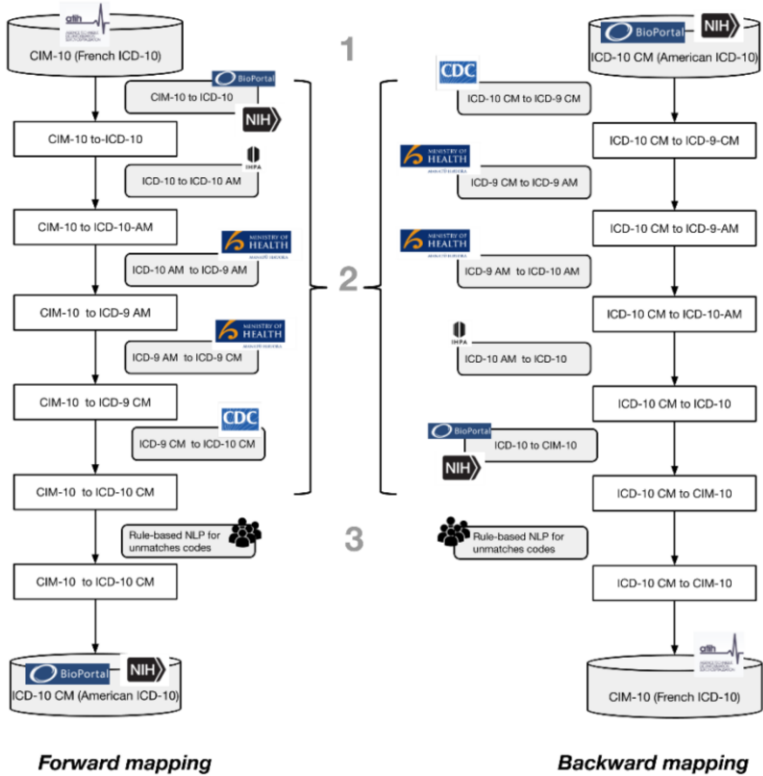


Figure 1. Forward and backward mapping methods

Since the automated mapping was based on official mapping files, we did not review those matched codes. However, we reviewed manually all NLP-based matches before confirming their match or un-match status. Uncertain pairs were reviewed again by two other medical investigators.

3. Results

The ICD-10 CM included 91,737 codes and the French ICD-10 included 39,928 codes, including the 3-characters codes parents. The ICD-10 CM had 9835 (11%) of 4-characters codes while the ICD-10 had 12,345 (31%) of 4-character codes.

Among the 39,928 codes of the CIM-10, 8,477 (21.2%) were exact matches to the ICD-10 CM and 29,131 (73 %) were partial matches. 264 of those partial matches were based on the NLP-based step. There were 2,320 (5.8%) missing codes.

Among the 91,737 codes of the ICD-10 CM, 9,082 (9.9%) were exact matches and 82,655 (90.1 %) partial matches with no unmatched codes. 226 of those partial matches were based on the NLP-based step. There were no missing codes.

All the NLP-based matches were true positives with no false negatives. Overall, the backward mapping was 94.2%, while the forward mapping was 100%. (Table 1)

Table 1. Characteristics of all matches after the forward and backward mappings

Mapping	Exact Match	Partial Match*				No match
		>4-characters code to one 4-character code	>4-characters code to more than one 4-characters code	4-characters code to more than one 4-characters code	4-characters codes (or more) to the 3-characters parent	
ICD-10 to ICD-10 CM	8,477 (21.2%)	4,832 (12.1%)	23,345 (58.5%)	932 (2.3%)	22 (0.1%)	2,320 (5.8%)
ICD-10CM to ICD-10	9,082 (9.9%)	41,518 (45.3%)	33,893 (36.9%)	1,797 (2%)	5,447 (5.9%)	0 (0%)

*The partial match includes the codes matched using the Natural Language Processing (NLP) step

4. Discussion

Our method showed a very high matching score, especially regarding the backward mapping (ICD-10 CM to ICD-10) with 100% of codes matched. However, the exact match was far better in the forward mapping (21% versus 10% in the backward). The manual evaluation confirmed the accuracy of the rules-based NLP algorithm.

Our study has several limitations. First, we only tested the method with two languages (French and English). However, we only had native speaking experts in French and English available for the manual review and most countries use either the ICD-10 CM, or a fairly close version of the WHO ICD-10 for reimbursement and billing[1]. Therefore, we knew that if our mapping process was accurate it would be relatively easy to adapt it to other languages. Second, our NLP was rather basic because we wanted first and foremost to use already existing reference files. Previous studies[11,12] showed that NLP based on modern machine-learning methods is the most pertinent method to match a diagnostic to an ICD-10 code or to translate a terminology to another language, but it is a very specific field with few experts, especially outside of the English language. The idea here was to propose an alternative when NLP specialists are not available to implement automatic translation and/or matching algorithms. Finally, most of our matches are partial. A majority of exact matches would have been ideal,

especially since the ICD-10 is not always very precise regarding some diagnoses[13], but since the WHO-ICD-10 and the ICD-10-CM have a vastly different number of codes[3], this outcome was predictable.

Manual mapping for multilingual ontology alignment is still the gold standard today, but it requires several experts of the healthcare and ontology fields and is a very time-consuming work that can take years[14]. This method cannot make the same claims of precision as official manual mapping files, but it could become a fairly quick and reliable process for international studies based on secondary reuse of EHR and billing data (including legacy data coded in ICD-9 CM, thanks to the GEM files from the NCHS[8]) without any ontology or translation experts.

5. Conclusion

Our study demonstrated that semi-automated mapping based on reference mapping files (in standard format) and basic NLP could be considered for secondary reuse of EHR and billing data from different countries when there are no existing reference files. Next, we would like to replicate this study with ICD-10 in other languages and use more automated NLP resources, like the Google Translate API, to confirm the accuracy of this method. This work could also serve as a basis for semi-automated mapping of the ICD-10 to the ICD-11, once the official ICD-10 to ICD-11 mapping files are available.

References

- [1] WHO | International Classification of Diseases (ICD) Information Sheet, *WHO*. (n.d.). <https://www.who.int/classifications/icd/factsheet/en/>.
- [2] ATIH : Agence technique de l'information sur l'hospitalisation, (n.d.). <http://www.atih.sante.fr/>.
- [3] L. Manchikanti, A.D. Kaye, V. Singh, and M.V. Boswell, The Tragedy of the Implementation of ICD-10-CM as ICD-10: Is the Cart Before the Horse or Is There a Tragic Paradox of Misinformation and Ignorance?, *Pain Physician*. **18** (2015) E485-495.
- [4] US National Library of Medicine. Unified Medical Language System (UMLS): Metathesaurus, (n.d.). https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html.
- [5] C.T. Dos Santos, P. Quaresma, and R. Vieira, An API for multilingual ontology matching, in: 2010.
- [6] P.L. Whetzel, N.F. Noy, N.H. Shah, P.R. Alexander, C. Nyulas, T. Tudorache, and M.A. Musen, BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications, *Nucleic Acids Res.* **39** (2011) W541-545.
- [7] A. Doan, J. Madhavan, P. Domingos, and A. Halevy, Ontology matching: A machine learning approach, in: *Handb. Ontol.*, Springer, 2004: pp. 385–403.
- [8] ICD - ICD-10-CM - International Classification of Diseases, Tenth Revision, Clinical Modification, (2019). <https://www.cdc.gov/nchs/icd/icd10cm.htm>.
- [9] Mapping between ICD-10 and ICD-9 | Ministry of Health NZ, (n.d.). <http://www.health.govt.nz/health-statistics/data-references/mapping-tools/mapping-between-icd-10-and-icd-9>.
- [10] Australian Consortium for Classification Development, The international statistical classification of diseases and related health problems, tenth revision, Australian modification (ICD-10-AM/ACHI/ACS): Mapping files, (n.d.). <https://ace.ihsa.gov.au/Downloads.aspx>.
- [11] N.D. Hailu, K.B. Cohen, and L.E. Hunter, Ontology translation: A case study on translating the Gene Ontology from English to German, *Nat. Lang. Process. Inf. Syst. Int. Conf. Appl. Nat. Lang. Inf. Syst. NLDB Revis. Pap. Int. Conf. Appl. Nat. Lang. Info.* **8455** (2014) 33–38.
- [12] A. Atutxa, A.D. de Ilaraza, K. Gojenola, M. Oronoz, and O. Perez-de-Viñaspre, Interpretable deep learning to map diagnostic texts to ICD-10 codes, *Int. J. Med. Inf.* **129** (2019) 49–59.
- [13] Y.M. Mesfin, A.C. Cheng, A.H.L. Tran, and J. Buttery, Positive predictive value of ICD-10 codes to detect anaphylaxis due to vaccination: A validation study, *Pharmacoepidemiol. Drug Saf.* (2019).
- [14] ICD-11 Content - Content Development Roadmap - SNOMED Confluence, (n.d.). <https://confluence.ihtsdotools.org/display/CDR/ICD-11+Content>.

An Automatic Image Collection System for Multicenter Clinical Studies

Toshihiro TAKEDA ^{a,1}, Shirou MANABE ^{a,b}, Atsushi HATTORI ^b,
Junji YAMAGUCHI ^b, Shozo KONISHI ^a, Yuichiro YAMAMOTO ^b,
Daiyo TAKAHASHI ^b, and Yasushi MATSUMURA ^{a,b}

^aDepartment of Medical Informatics, Osaka University Graduate School of Medicine

^bMKS Ltd

Abstract. The acquisition of medical images from multiple medical institutions has become important for high-quality clinical studies. In recent years, electronic data submission has enabled the transmission of image data to independent institutions more quickly and easily than before. However, the selection, anonymization, and transmission of medical images still require human resources in the form of clinical research collaborators. In this study, we developed an image collection system that works with the electronic data capture (EDC) system. In this image collection system, medical images are selected based on EDC input information, patient ID is anonymized to a subject ID issued by the EDC, and the selected anonymized images are transferred to the research institute without human intervention. In the research institute, clinical information registered by the EDC and clinical images collected by the image collection system are managed by the same subject ID and can be used for clinical studies. In October 2019, our image collection system was introduced to 13 medical institutions and has now begun collecting medical images from the in-hospital picture archiving and communication system (PACS) of those institutions.

Keywords. Multicenter clinical study, Electronic data submission, Medical image

1. Introduction

Assessments of medical images are often needed in multicenter clinical studies. To ensure the quality of image analyses, imaging core labs that collect medical images taken at multiple medical institutions in one place and analyze them in a defined way are needed [1, 2]. Regarding image analyses by artificial intelligence (AI), a large amount of medical images are required [3, 4]. For this reason, a system that supports efficient image collection from independent institutions is required.

For the protection of patient information, medical images used for clinical research must be anonymized at each medical institution. Medical images stored in the digital imaging and communications in medicine (DICOM) standard can be anonymized by rewriting the DICOM tag information. However, in a non-DICOM image, patient information is described in the image, and it is necessary to erase patient information by image processing for anonymization.

¹ Corresponding Author, Toshihiro Takeda, Department of Medical Informatics, Osaka University Graduate School of Medicine, 2-2 Yamada-oka Suita Osaka 565-0871 Japan, E-mail: ttakeda@hp-info.med.osaka-u.ac.jp

Clinical information must be collected in addition to images for subsequent analyses. Such clinical information is often collected via an electronic data capture (EDC) system [5, 6]. In EDC system, anonymization is usually performed by replacing the patient ID with the subject ID. For efficient data analyses in a research institute, the anonymization of clinical images is also performed using the same subject ID.

In recent years, electronic data submission has enabled the transmission of image data to independent institutions more quickly and easily than before [7]. However, research collaborators in each institution still need to select the medical images, anonymize the patient information, and perform the electronic transfer itself. When it is difficult to replace the patient ID with the subject ID, a correspondence table must be created between the image anonymization ID and the subject ID and sent to the relevant research institute. These tasks require substantial effort from research collaborators, who are already busy with their clinical work.

We developed an EDC system that collects clinical information in cooperation with electronic medical records (EMRs) [8]. The doctors in each collaborating institution enter clinical information via template of the EDC system which can retrieve the data recorded in the EMR. The EDC system outputs narrative text of input data to the progress note of the EMR. The EDC system also maps the input data to the operational data model (ODM) developed by Clinical Data Interchange Standard Consortium (CDISC) and sends to the research institute. The EDC site server anonymizes the patient ID and issues a subject ID, maintaining a correspondence table of the patient and subject IDs.

In the present study, we developed an image collection system that works with our EDC system. In this image collection system, medical images are selected based on EDC input information, a patient ID is anonymized to a subject ID issued by the EDC system and the selected anonymized medical images are transfer to the research institutes.

2. System Overview

The image collection system is divided into image acquisition from the picture archiving and communication system (PACS) and that from non-DICOM server. The EDC site server and the image acquisition gateway were set up in each collaborating institution (Figure 1), while the EDC center server and image-receiving server were set up in the research institute. A virtual private network (VPN) internet connection was used for connections between the research institute and each medical institution.

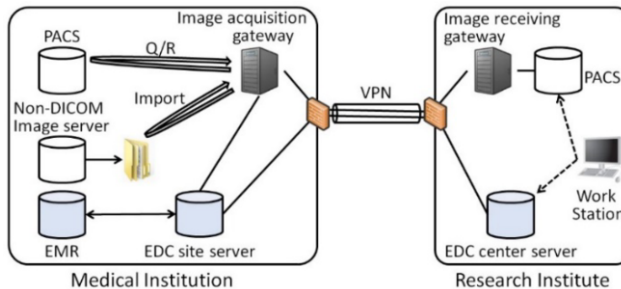


Figure 1. System overview

3. Collection of images stored in the DICOM standard in prospective studies

3.1. Image acquisition and anonymization by the image acquisition gateway

The image acquisition gateway asks the EDC site server about image information to be acquired. This information includes the patient ID, observation date, modality, and observation area. The subject ID is also obtained at this time. Next, the gateway searches the PACS for the target images using a query and retrieve (Q/R) connection. The patient ID described in the DICOM tag of the acquired images is then anonymized by replacing it with the subject ID issued by the EDC system. Finally, the gateway transfers the anonymized images to the image receiving server in the research institute.

3.2. Transmission of information from the EDC to the image acquisition gateway

Our EDC system is designed to enter clinical information using templates for each event. For the image to be collected, the item of “observation date” is set in the template. When a study using collected images is registered to the EDC, the EDC outputs the ODM files of the input template data to the predetermined folder at a time fixed by the scheduler (Figure 2). The output ODM file name is composed of the “patient ID”, “Study OID”, “Event OID”, “ODM file output date and time”. “Subject ID”, and input data, including the “observation date”, are described in the ODM file.

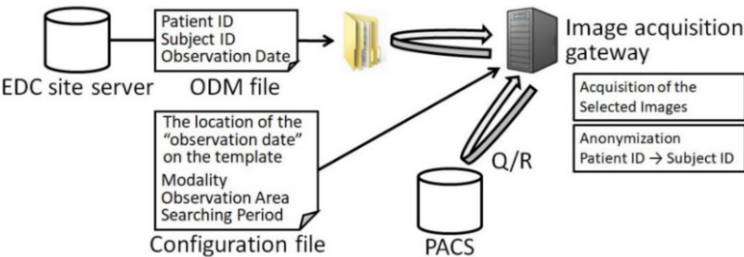


Figure 2. Transmission of information from the EDC to the image acquisition gateway

We defined the configuration file for collecting images beforehand and registered it to the image acquisition gateway. In the configuration file, information about the location of the “observation date” described on the template are defined. The location is identified by the “Study Code”, “Study Event OID”, “Form OID”, “Item Group OID”, and “Item Data OID”. In the configuration file, the “modality”, “observation area”, and “search period” based on the “observation date” are also defined. The image acquisition gateway acquires search conditions for images to be acquired from the ODM file output by the EDC system and the configuration file registered beforehand and searches for images stored in the in-hospital PACS. Figure 3 shows an example of a configuration file for extracting images from the in-hospital PACS.

Study Code	Study Name	Image Identification Name	Modality	Observation Area	Study Event OID	Form OID	Item Group OID	Item Data OID	Search Period		Comment	Destination	Status 0: waiting 1: stopped
									Pre	Post			
VS4	VS4 study	Preoperative CT	CT	CHEST	SE PREOP	F.CT	IG.CT	ID.CT DAT	6	1	CT images taken just before surgery	A00	0
VS4	VS4 study	Postoperative CT (1M)	CT	CHEST	SE 1M	F.CT	IG.CT	ID.CT DAT	3	4	CT images taken 1 month after surgery	A00	0

Figure 3. The EDC template and its configuration file for collecting images

4. Collection of images stored in the non-DICOM standard

For the collection of non-DICOM images, the research collaborators must select the images and output them to a specific folder (Figure 1). Next, the patient information described in the selected images is masked using the image acquisition software program. This program can set a plurality of masking patterns in advance, and one of them is set as the main masking pattern for each project. After masking the patient information, the images are converted to the DICOM format, and the patient IDs in the DICOM tag are anonymized to the subject IDs issued by the EDC system. The images are then transferred to the image-receiving server placed in the research institute.

The application can collect the image information from the image folder and file name. There are two patterns for collecting image information: (1) a pattern that sets a delimiter and corresponds to variable-length information, and (2) a pattern that does not set a delimiter and corresponds to fixed-length information. Figure 4 shows an example of the image information collected from a folder and file name.

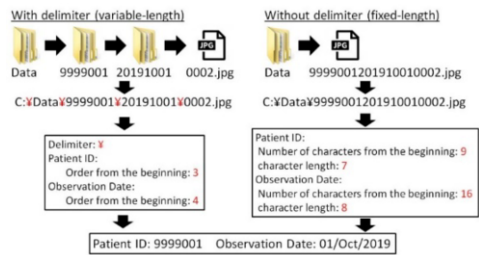


Figure 4. Collection of the non-DICOM image information

5. Collection of images stored in the DICOM standard in retrospective studies

Images obtained in the past may need to be collected in bulk, such as for image analyses using AI. However, as the PACS is designed to be used on a per-patient basis, the selection and download of images must be performed by launching the PACS viewer for each patient. The image collection system can batch download medical images, anonymize the patient ID to the subject ID, and send the images to the research institute according to the acquisition conditions defined in the csv format file.

6. Introduction status of our image collection system

On October 1, 2019, our image collection system was introduced to 13 medical institution, and medical images began to be collected from these centers' in-hospital PACS, which had been developed by 4 PACS manufacturers.

7. Discussion

As discrepancies may occur between the image analysis results assessed by an imaging core lab and those obtained by study sites, it is necessary to collect images at a single institute to maintain the quality of clinical research involving image analyses [1].

Although the submission of imaging data in an electronic format has begun to spread, the tasks of selecting, anonymizing, and transferring data still pose a substantial burden to research collaborators. This results in a limited number of medical institutions cooperating in image collection. Our image collection system selects the images to be collected, anonymizes the patient information, and transfers them to the research institute. Since the images to be collected are determined on a study-by-study basis, the modality and observation area of the images are entered into the configuration file and registered in the image acquisition gateway, so that research collaborators only need to add the observation date of the images to the EDC template. This image collection system may facilitate multi-center clinical studies.

For images stored in the DICOM format, patient information is only described in the DICOM tag. Therefore, automatic anonymization can be easily achieved. In contrast, the images stored in the non-DICOM format may include sensitive patient information, so image processing is necessary for anonymization. To ensure reliable anonymization, automatic anonymization should not be performed. However, setting the masking area and performing the automatic acquisition of image information from the folder and file names will assist in collecting non-DICOM images.

8. Conclusion

We developed an image collection system that automatically selects, anonymizes, and transfers medical images based on the EDC input information.

Acknowledgments

This work was supported by Council for Science, Technology and Innovation, cross-ministerial Strategic Innovation Promotion Program, "Innovative AI Hospital System" (Funding Agency: National Institute of Biomedical Innovation, Health and Nutrition).

References

- [1] Fahed R, Ben Maacha M, Ducroux C, et al. Agreement between core laboratory and study investigators for imaging scores in a thrombectomy trial. *J Neurointerv Surg*. 10 (2018) e30.
- [2] Miller JM1, Rochitte CE, Dewey M, et al. Diagnostic performance of coronary angiography by 64-row CT. *N Engl J Med*. 359 (2008) 2324-36.
- [3] Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng*. 21 (2017) 221-248.
- [4] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 42 (2017) 60-88.
- [5] Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 42 (2009) 377-81.
- [6] Pan HY, Shaitelman SF, Perkins GH, et al. Implementing a Real-Time Electronic Data Capture System to Improve Clinical Documentation in Radiation Oncology. *J Am Coll Radiol*. 13 (2016) 401-7.
- [7] Goldmacher GV, Raunig D. The imaging core lab perspective on glioblastoma imaging and response assessment in clinical trials. *Neuro Oncol*. 16 (2014) 48-50.
- [8] Matsumura Y, Hattori A, Manabe S, et al. Case Report Form Reporter: A Key Component for the Integration of Electronic Medical Records and the Electronic Data Capture System. *Stud Health Technol Inform*. 245 (2017) 516-520.

An Evolutionary Approach to the Annotation of Discharge Summaries

Christina LOHR^{1ab}, Luise MODERSOHN^{ab}, Johannes HELLRICH^{ab}, Tobias KOLDITZ^{ab}, Udo HAHN^{ab}

^a*Jena University Language & Information Engineering (JULIE) Lab Friedrich-Schiller-Universität Jena, Germany*

^b*SMITH Consortium of the German Medical Informatics Initiative*

Abstract. We here describe the evolution of annotation guidelines for major clinical named entities, namely *Diagnosis*, *Findings* and *Symptoms*, on a corpus of approximately 1,000 German discharge letters. Due to their intrinsic opaqueness and complexity, clinical annotation tasks require continuous guideline tuning, beginning from the initial definition of crucial entities and the subsequent iterative evolution of guidelines based on empirical evidence. We describe rationales for adaptation, with focus on several metrical criteria and task-centered clinical constraints

Keywords. Clinical text corpus, German discharge summaries, Annotation guideline

1. Introduction

The annotation of clinical documents is a challenging task due to the intrinsic opaqueness and complexity of the clinical domain. Clinical experts do not necessarily agree on fundamental judgments relevant for annotation decisions, e.g., the distinction between symptoms, diagnoses and other medically relevant phenomena. Reports of medical annotation campaigns often disregard this lack of conclusiveness insofar as the painstaking adaptive processes underlying the vernier adjustment of annotation guidelines are hidden from the scientific community. Most often only the final outcome of consensus-seeking processes is made publicly available. Annotated clinical text corpora focusing on diagnoses came up with clinical shared tasks, e.g., I2B2 [1], CLEF eHEALTH [2], and SEMEVAL [3]. The annotation of major clinical entities is typically considered as a very hard decision task, with agreement scores (*F*-scores, κ - or α -values) usually in the range between 0.7 and 0.8, even after several modification rounds for annotation guidelines (see Table 1 for an overview of previous annotation campaigns related to clinical documents).

We here want to shed light on the impact of different annotation policies and propose robust metrical decision criteria in order to find out which kinds of adaptive steps have particular impact on annotators' (dis)agreement and how such factors can be controlled by empirical evidence. We illustrate an evolutionary approach to the development of

¹ Corresponding Author: email: christina.lohr@uni-jena.de

annotation guidelines for major clinical named entities, namely *Diagnoses*, *Findings* and *Symptoms*, on a corpus of roughly 1,000 German discharge summaries.

Table 1. Overview of clinical annotation studies, incl. annotated entities and agreement scores (F -score, Cohen’s κ (κ_c), Siegel and Castellan’s κ (κ_{sc}) or Krippendorff’s α)

Authors	Data	Annotated Entities	Agreement
English language text corpora			
Wang et al. [4]	300 clinical notes	SNOMED-CT concepts	$F = 0.88$
Savova et al. [5]	273 clinical notes	UMLS semantic groups	$\kappa_c = 0.727$
Doğan et al. [6]	793 PubMed abstracts	diseases	$F \approx 0.9$
Albright et al. [7]	13K pathol. report sentences	UMLS semantic groups	$F \in [0.7, 0.75]$
Hahn et al. [8]	200 Medline abstracts	diseases, symptoms, anatomy	$\kappa_{sc} \in [0.7, 0.8]$
Patel et al. [9]	5K different documents	UMLS semantic groups	$\kappa_c \in [0.68, 0.97]$
Multilingual text corpora (including English and German language)			
Miñarro-Giménez et al. [10]	60 short texts	terminology terms SNOMED-CT and UMLS	$\alpha = [0.4, 0.9]$
Kors et al. [11]	800 Medline titles, 500 drug labels, 150 patent claims	biomedical concepts based on UMLS semantic groups	$F = 0.79$
Non-English language text corpora			
Toepfer et al. [12]	140 German ECG reports	fine grained information extraction	$F = 0.955$
Skeppstedt et al. [13]	1,148 Swedish EPR reports	disorders, findings, pharma	$F \in [0.69, 0.88]$
Campillos et al. [14]	500 French clinical notes	entities derived from UMLS semantic groups	$F = 0.793$
He et al. [15]	138 Chinese documents		$F = 0.992$

2. Methods

Data and Annotation Setup. This work was carried out on the Jena part of the 3000PA text corpus [16]—1,106 German discharge summaries from the Jena University Hospital’s information system given the following constraints: patients had stayed for at least five days on a ward for internal medicine or in an intensive care unit between 2010 and 2015, and were already deceased. These criteria were approved by the local ethics committee (4639-12/15). The corpus consists of approx. 170K sentences and 1.5M tokens. The documents were annotated (using the *Brat Rapid Annotation Tool*)² by eight medical students who all had passed their first medical licensing exam. We ran through four iterations of guideline revisions, tool configuration, team instruction, annotation work, agreement computation, error analysis and subsequent team discussion. This process was supervised by one annotation manager. In each iteration, 5 to 10 documents and 50 in the final round were annotated by each student (for agreement values see Table 2).

Evaluation Measures. A challenging problem for medical annotations is posed by determining the proper text span of named entities. We thus calculate inter-annotator agreement (IAA) by using the pair-wise average F -score [17] of instances and tokens. An *instance* is a single composite annotation unit that consists of one or more tokens (e.g., “renal insufficiency” denotes an instance with two tokens, “renal” and “insufficiency”). Instance agreement is a strong criterion, as annotations with only slightly differing spans

² <https://brat.nlplab.org/>

count as non-matching. Token-based agreement describes the overlap of individual tokens for annotations of the same instance type³.

Table 2. Inter-annotator agreement (IAA; pair-wise average F-score incl. standard deviation (σ)) of training iterations and the final annotation round (incl. the number of documents being used per iteration (in brackets)), for each document over all eight annotators w.r.t. instances (inst.) and tokens (tok.)

	Iteration 1 (10)		Iteration 2 (5)		Iteration 3 (10)		Iteration 4 (10)		Final (50)	
Category	inst.	tok.	inst.	tok.	inst.	tok.	inst.	tok.	inst.	tok.
Diagnosis	.504	.633	.534	.630	.593	.705	.614	.688	0.637	0.758
Findings	.354	.545	.698	.896	.682	.887	.628	.864	0.684	0.860
Symptoms	.583	.617	.441	.657	.424	.729	.532	.669	0.611	0.718
Anatom. Loc.	.455	.624	.735	.876						
Procedures	.252	.301	.527	.580						
Avg. F-score	.410	.563	.671	.832	.653	.873	.621	.837	.648	.839
σ (F-score)	.079	.076	.056	.030	.064	.027	.065	.036	.056	.032

Annotation Schema. We started our annotation campaign with five types of medical entities: *Diagnosis* as *diseases* as listed in the ICD-10 (excluding symptom chapter ‘R’); *Findings* as the summary of all information about a patient’s medical status reported by medical staff based on the results from physical examination or the use of medical devices (e.g., x-ray); *Symptoms* as the summary of all health affecting information and complaints reported by the patients or their relatives; *Anatomical Location* as the summary of all anatomical parts of the body excluding body fluids and body secretions; *Procedures* as the summary of all medical interventions like surgery, non-surgical therapeutic measures, etc. The administration of (new) medication belongs to the latter category, but was excluded here since medication had already been annotated elsewhere[16].

Iterative Process. The average agreement value (*F*-score) of the first iteration reached .410 on instances and .563 on tokens which is clearly below expectations. During the second iteration *Anatomical Locations* were automatically pre-annotated based on the German part of the *Foundational Model of Anatomy (FMA)* of the UMLS using the JUFIT tool. This step increased the IAA for *Anatomical Location* up to .735 on instances and .876 on tokens. Still, it was removed from further annotation rounds because of its overlap with mentions of *Diagnoses*, *Findings* or *Symptoms*. Also discussions with the annotators pointed at a painfully large number of single decisions and double annotations to be made resulting in an average annotation time of one hour per document. In particular, distinguishing *Procedures* from *Medications* (e.g., “long-term oxygen therapy”) and *Diagnoses* was considered a difficult task—a previous intervention can be coded as *Procedures* and *Diagnoses* (e.g., Z98.8—Other specified postsurgical states). As a clear criterion to minimize the overlap between *Diagnoses* and *Procedures* could not be found we eliminated *Procedures* from the current annotation project and will deal with it separately. Following the approach of Hahn et al. [8] for annotating long and short text spans, we required the attribute *Complexity* to be tagged for all entities containing more than one piece of annotation-relevant information (e.g., “Pupils middle and isocor”). *Diagnosis*, *Symptoms* and *Findings* were assigned two attributes: *Time* with the value set *previous*, *recurrent*, *uncertain* and *Modality* with the value set *suspected*, *excluded*, *uncertain*. Negations or vague descriptions were defined to be part of the long annotation span.

³ We used TREE TAGGER, accessible via <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

Baseline Classifier. We trained a (CRF-based) JCORE pipeline for named entity recognition of *Diagnosis*, *Findings* and *Symptoms*; evaluation with 10-fold cross-validation.

Table 3. Instances and tokens of entity types *Diagnosis*, *Findings*, *Symptoms*, with the percentage distribution of their attributes *Complexity*, *Modality*, *Time*; baseline classifier (BC) results (avg. token distribution depending on all of the 1.863M tokens; all other percentage values depending on the total number of entity instances)

	Inst.	Tokens		Avg. span	Complexity	Modality			Time			BC
						sus.	exc.	unc.	rec.	unc.	pre.	
Entity type	freq.	freq.	%	tok.	%	%	%	%	%	%	%	F
<i>Diagnosis</i>	55K	123K	6.6	2.2	3.46	4.44	9.31	1.49	0.67	0.01	6.66	.475
<i>Findings</i>	150K	663K	35.6	4.4	17.15	1.24	17.36	1.40	0.13	0.02	1.81	.805
<i>Symptom</i>	8K	26K	1.4	3.3	14.21	0.10	22.16	0.13	0.96	–	1.57	.143
Σ Annotations	163K	812K	43.40	3.8	13.50	2.02	15.45	1.38	0.30	0.01	3.05	–

3. Results

The final average IAA values were .648 on instances and .839 on tokens (see Table 2). Table 3 illustrates the distribution of the different entities and their attributes. About one third of the entire corpus consists of *Findings*. *Diagnoses* added up to around 6.6%, whereas *Symptoms* were rare with less than 2%. Around 13% of all annotations were marked as *complex*; roughly every fifth instance of *Findings* and *Symptoms* contained more than one information unit. Compared to *Findings* and *Symptoms*, *Diagnoses* were more precisely described. 10–20% of all annotations were *negated*. The majority of negated entities were *Symptoms* and *Findings* ($\approx 20\%$). The attributes *suspected* and *uncertain* were rarely used, only about 5% of *Diagnosis* annotations were marked as *suspected*. Temporal annotations of *Findings* and *Symptoms* were hardly used, $\approx 7\%$ of *Diagnoses* were *previous*. The average *F*-score results of our (admittedly simple) CRF-based classifier are .805 for *Findings*, .475 for *Diagnosis* and .143 for *Symptoms*.

4. Discussion

The final IAA values (0.7–0.8) could be improved and are comparable to, if not higher than, annotation campaigns covering the listed entity types in languages other than German (cf. Table 1). Category overlap or hard to disentangle semantic interdependencies must be resolved for the sake of acceptable IAA values—either by generalizing category definitions, by stretching the granularity of annotation spans, by adding complexity markers (thus delegating finer distinctions to future annotation studies), or, in the worst case, by eliminating sloppy named entity types from the current annotation process.

5. Conclusion

We presented an evolutionary approach to an annotation scheme for *Diagnoses*, *Findings* and *Symptoms* and its application to German discharge summaries. We struggled with

criteria that control the number of iterations and the directions of adaptations. Generalizing from the our annotation task, we found that IAA is a solid metrical indicator for task complexity or inconclusive annotation guidelines. Both parameters have to be tuned in case of too low IAA values. Annotation time is a metrical indicator for task complexity.

Acknowledgements Work supported by BMBF within the SMITH project under grant 01ZZ1803G and DFG under grant HA 2079/8-1 within the STAKI2B2 project. We thank all annotators, as well as André Scherag, Danny Ammon, and all members of the Data Integration Center of the Jena University Hospital.

References

- [1] Uzuner O, South BR, Shen S, and DuVall SL, “2010 i2B2/VA CHALLENGE on concepts, assertions, and relations in clinical text,” *JAMIA*, vol. 18, no. 5, pp. 552–556, 2011.
- [2] Suominen H, Zhou L, Hanlen L, and Ferraro G, “Benchmarking clinical speech recognition and information extraction: new data, methods, and evaluations,” *JMI*, vol. 3, no. 2, p. e19, 2015.
- [3] Elhadad N, Pradhan SS, Lipsky Gorman S, Manandhar S, Chapman WW, and Savova GK, “SemEval-2015 Task 14: Analysis of Clinical Text,” in *SemEval 2015*, pp. 303–310, 2015.
- [4] Wang Y, “Annotating and recognising named entities in clinical notes,” in *Proceedings of the Student Research Workshop @ ACL-IJCNLP 2009*, pp. 18–26, 2009.
- [5] Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, and Chute CG, “Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications,” *JAMIA*, vol. 17, no. 5, pp. 507–513, 2010.
- [6] Doğan RI, Leaman R, and Lu Z, “NCBI disease corpus: a resource for disease name recognition and concept normalization,” *Journal of Biomedical Informatics*, vol. 47, pp. 1–10, 2014.
- [7] Albright D, Lanfranchi A, Fredriksen A, Styler I, William F, Warner C, Hwang JD, Choi JD, Dligach D, Nielsen RD, Martin J, Ward W, Palmer M, and Savova GK, “Towards comprehensive syntactic and semantic annotations of the clinical narrative,” *JAMIA*, vol. 20, pp. 922–930, 012013.
- [8] Hahn U, Beisswanger E, Buyko E., Faessler E, Traumüller J, Schröder S, and Hornbostel K, “Iterative refinement and quality checking of annotation guidelines: How to deal effectively with semantically sloppy named entity types, such as pathological phenomena,” in *LREC 2012*, pp. 3881–3885, 2012.
- [9] Patel P, Davey D, Panchal V, and Pathak P, “Annotation of a large clinical entity corpus,” in *EMNLP 2018*, pp. 2033–2042, 2018.
- [10] Miñarro-Giménez JA, Cornet R, Jaulent MC, Dewenter H, Thun S, Rosenbeck Gøeg K, Karlsson D, and Schulz S, “Quantitative analysis of manual annotation of clinical text samples,” *International Journal of Medical Informatics*, vol. 123, pp. 37–48, 2019.
- [11] Kors JA, Clematide S, Akhondi SA, van Mulligen EM, and Rebholz-Schuhmann D, “A multilingual gold-standard corpus for biomedical concept recognition: the MANTRA GSC,” *Journal of the American Medical Informatics Association*, vol. 22, no. 5, pp. 948–956, 2015.
- [12] Toepfer M, Corovic H, Fette G, Klügl P, Störk S, and Puppe F, “Fine-grained information extraction from German transthoracic echocardiography reports,” *BMC MIDM*, vol. 15, p. #91, 2015.
- [13] Skeppstedt M, Kvist M, Nilsson GH, and Dalianis H, “Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: An annotation and machine learning study,” *Journal of Biomedical Informatics*, vol. 49, pp. 148–158, June 2014.
- [14] Campillos L, Deléger L, Grouin C, Hamon T, Ligozat AL, and Nèvéol A, “A French clinical corpus with comprehensive semantic annotations: development of the Medical Entity and Relation LIMSI annotated Text corpus (MERLOT),” *Language Resources and Evaluation*, vol. 52, pp. 571–601, 2018.
- [15] He B, Dong B, Guan Y, Yang J, Jiang Z, Yu Q, Cheng J, and Qu C, “Building a comprehensive syntactic and semantic corpus of Chinese clinical texts,” *JBI*, vol. 69, pp. 203–217, May 2017.
- [16] Hahn U, Matthias F, Lohr C, and Löffler M, “3000PA: Towards a national reference corpus of German clinical language,” in *MIE 2018*, no. 247, pp. 26–30.
- [17] Hripcsak G and Rothschild AS, “Agreement, the f-measure, and reliability in information retrieval,” *Journal of the American Medical Informatics Association*, vol. 12, no. 3, pp. 296–298, 2005.

Application of Radiomics in Vesselness Analysis of CT Angiography Images of Stroke Patients

Daria DOLOTOVA^{a,1}, Ivan ARKHIPOV^a, Evgenia BLAGOSKLONOVA^a, Victoria DONITOVA^b, Tatiana BARMINA^c, Faat SHARIFULLIN^c, Boris KOBRINSKII^b and Andrey GAVRILOV^{a,d}

^a*Gammamed-Soft, Ltd, Moscow, Russia*

^b*Federal Research Center "Computer Science and Control" of RAS, Moscow, Russia*

^c*N.V.Sklifosovskiy Scientific Research Institute of Emergency Care, Moscow, Russia*

^d*Lomonosov Moscow State University, Moscow, Russia*

Abstract. Development of vascular collaterals in a lesion area is one of the key factors that determine not only the choice of treatment for ischemic stroke (IS) patients, but also outcome and therapy effectiveness. The main method for examining the vessels' ramification is CT angiography (CTA). CTA analysis may be improved by incorporating filters designed to extract more features about vessels and quantify their level of development. This work suggests the usage of radiomics methods in the analysis of vesselness measure calculated from CTA images. Vesselness measurement is based on the analysis of the Hessian matrix with a few modifications dictated by practical aspects of this issue. The developed algorithm was implemented as a filter that generates a new 3D image, every voxel of which has the probability of belonging to a vessel-like structure. Further analysis of the distribution of vesselness in the lesion area and in the intact contralateral area was conducted with the methods from the open library PyRadiomics. A set of radiomics features was calculated. Preliminary analysis on a sample of 30 IS patients showed the presence of significant differences between afflicted and intact hemispheres.

Keywords. Collaterals, imaging biomarkers, ischemia, tomography

1. Introduction

Ischemic stroke (IS) is one of the leading causes of death and disability in the world [1]. The outcome of IS depends not only on the timeliness of the reperfusion therapy but also on the intensity of collateral flow [2]. The absence of well-developed collaterals leads to an increase in the volume of necrotized tissue and to a decrease in the effectiveness of treatment [3]. In daily practice, CT angiography (CTA) is used as the main method for visualizing of vascular anatomy. However, it is very common among radiologists to analyse CTA images manually, resulting in subjectivity of the assessment. Cutting-edge methods of image processing, including vesselness calculation algorithms, may be of help in addressing this problem, by providing an automatic measure of collateral status.

Vesselness calculation involves assessment of the probability of each voxel belonging to a vessel-like structure [4]. According to literature, there is a possibility of using vesselness to resolve some clinical challenges in treatment of IS patients [5-7]. Currently, most of studies include analysis of variables based only on the volume of vessel-like structures, but not on their spatial behavior and other characteristics. At the same time, one of the most perspective directions of modern medicine is the analysis of radiological images, specifically radiomics [8-10]. It allows extracting numerous features including those that describe shape and texture from radiological images. So, in the line with the above, the aim of the presented work was the application of radiomics methods in the analysis of vascular behavior calculated from the CTA images of the patients with IS.

2. Methods

This pilot study included data of 30 patients (74 years old (IQR 57-80), 18 males, 12 females) admitted to the N.V. Sklifosovskiy Scientific Research Institute of Emergency Care with IS in the territory of the middle cerebral artery (up to 5 hours from stroke onset). The examination of the patients was performed with multi-slice CT-scanner (GE Medical Systems), with pixel spacing of 0.44 x 0.44 mm, and slice thickness of 0.5 mm. We used CTA images obtained on the admission of a patient to the hospital, and images from non-enhanced CT conducted on the second day after IS onset, when the ischemic region was already visible as a hyperdense area. The last ones were used for contouring of the regions of interest (ROI) that was conducted by two radiologists (Figure 1, A). Each of them had more than 20 years of experience.

Brain normalisation was conducted to build a symmetrical area and to transfer both areas (with damaged and intact brain tissues) onto the CTA series (Figure 1, B and D). It was performed with the set of instruments described in the Advanced normalisation tools (ANTS) [11] that allowed coregistration of a brain to a template image using linear affine and warp transformations.

Then, CTA images were processed through the vesselness filter (Figure 1, C). The vesselness calculation was based on the following key concept: if we look at every voxel of a 3D image and evaluate the function of the derivative of intensity along three orthonormal directions, then it is possible to establish criteria for a voxel belonging to a tubular structure. The mathematical representation of this is the Hessian matrix calculation with second-order partial derivatives, and the assessment of its three eigenvalues λ_1 , λ_2 , λ_3 sorted in ascending order [4]. This method successfully identifies vessels, but it causes a decline of vesselness value in vessel bifurcation points. To address this issue, [7] proposed using λ_p as a regularized value, which is computed depending on the maximal value of λ_3 .

Based on the latter, we developed a vesselness filter, but made some changes to adaptation to further clinical research work. First, as the original CT images are noisy, it was necessary to smooth them to make them better suited for vessel analysis. The minimisation of noise is determined by the size of vessels that we are looking for and by the resolution of the CT scanner (0.7-1.4 mm). Thus, we applied the Gaussian filter with $\sigma=0.7$ mm. This allowed us to determine the Hessian matrix more robustly while keeping information about small vessels. Second, we adapted the approach proposed in [7] to λ_p regularization. λ_p depends on the maximal value of λ_3 throughout an image. In this case, even if we have a fairly smooth image with relatively meagre vascularity, we find a lot

of vessel-like structures. So, the results of vesselness analysis for different images and patients are not comparable. To avoid this, we decided to exclude the dynamic nature of λ_p and fix the maximum value of λ_3 at 400 HU/mm². The filter was implemented in our previously developed platform for medical images analysis [12].

The initial CTA images of 30 patients were processed through it, and new 3D images were generated. Each voxel has a vesselness value which measures the similarity of the voxel to a vessel-like structure. As the total sum of vessel-like voxels is very sensitive to the threshold of vesselness, we decided to implement the calculation of other features of the filtered images, describing other morphometric and distribution aspects of vesselness. The open library PyRadiomics was used to calculate 75 features of damaged and intact areas [13-14]. They included first-order statistics (FO) and second-order statistics which were based on analysis of the grey-level co-occurrence (GLCM), grey-level size zone (GLSZM), and grey-level run-length (GLRLM) matrices.

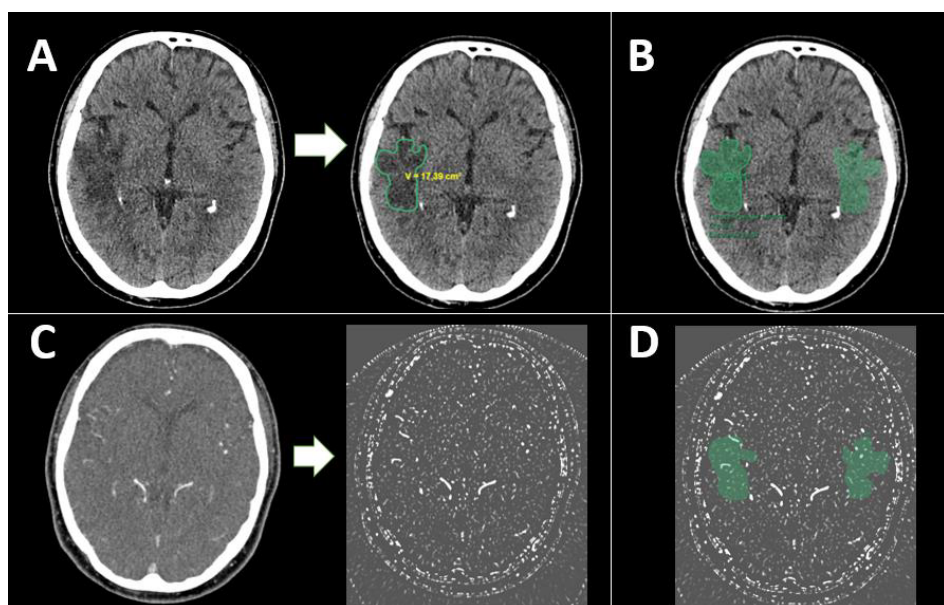


Figure 1. Scheme of the image processing for the calculation of radiomics features of vesselness (A – semiautomatic segmentation of ROI with damaged tissue on non-enhanced CT, B – automatic segmentation of ROI in the intact area, C – Application of vesselness filter on CTA images, D – copy of two ROIs to the filtered CTA images)

3. Results

Comparative analysis of 75 features in damaged and intact areas of 30 patients showed the presence of statistically significant differences in 57 parameters ($p < .05$, the Wilcoxon signed-rank test). For 48 of them these differences were highly significant ($p < .001$). We evaluated their relative change against the intact side to select the most informative features. It the confidence interval of a feature's relative change contained zero, the feature was excluded. Thus, the following list of features with the most pronounced and stable changes was formed (Table 1).

Table 1. The magnitude of the features' relative change in comparison with intact area.

Feature Group	Feature Name	Relative change	P-value
FO	Energy	↓ 79.3 (IQR 12.3-91.0) %	<<0.001
	Variance	↓ 79.6 (IQR 11.1-91.2) %	<<0.001
	Entropy	↓ 34.9 (IQR 11.1-91.2) %	<<0.001
GLCM	Contrast	↓ 81.4 (IQR 21.9-47.7) %	<<0.001
	Difference Variance	↓ 81.5 (IQR 21.0-91.0) %	<<0.001
	Joint Entropy	↓ 34.5 (IQR 23.3-47.3) %	<<0.001
GLRLM	High Grey Level Run Emphasis	↓ 72.0 (IQR 10.0-84.6) %	<<0.001
GLSZM	Large Area Emphasis	↑ 133.2 (IQR 10.9-228.1) %	<<0.001
	Size-Zone Non-Uniformity	↓ 71.9 (IQR 14.9-84.9) %	<<0.001
	Zone Variance	↑ 133.1 (IQR 10.9-228.0) %	<<0.001

As you can see in the table 1, the damaged areas show a decrease in energy and variance as a result of reduction in the number of vessel-like voxels. GLCM features indicated a decrease in the homogeneous patterns in the image. Changes in size-zone non-uniformity and zone variance of GLSZM reflect the fact that the distribution of zones with the same intensity level was modified in the damaged ROIs.

4. Discussion

One of the most effective ways to improve the quality of treatment for IS patients is to implement the advanced technologies of intellectual analysis and image processing into interpretation of the results of neuroimaging examinations. Assessment of the collateral status on the CTA allows predicting effectiveness of revascularization surgeries and long-term outcome of a clinical case, hence providing the decision support in the choosing of the treatment tactics for patients.

In this work we detected a large difference between damaged and intact areas in 10 features describing vesselness behavior. An interesting aspect of these preliminary results was that some of these features indirectly described the volume of vascular tree or a number of bright voxels (e.g., energy, variance, root mean squared), but others outlined the homogeneity of regions of interest (e.g., entropy, contrast of GLCM, size-zone non-uniformity and zone variance of GLSZM).

It is important to note that with the application of radiomics methods to the initial images of CTA it is difficult to explain whether the discovered differences in the features are caused by the changes in vessels or in the brain tissue. At the same time, the physiological interpretation of the images resulting from the use of the vesselness filter is simpler as all changes are only related to voxels of vessels.

Limitations to this study include its retrospective design and a small sample size. Also, we didn't have a control group to investigate the behavior of the features for healthy people. In the future work we will try to address these limitations.

5. Conclusion

In this paper, we presented a method of quantitative assessment of collaterals based on the analysis of vesselness measure. The suggested procedure includes the calculation of vesselness with an algorithm developed by Jerman et al, modified by using Gaussian filters and fixing the maximum value of λ_3 . The result of the application of the vesselness

filter is proposed to analyse with the radiomics methods, effectively investigating not only the volumetric characteristics of vesselness, but also the pattern of its distribution. Comparison analysis of extracted features of vesselness between the afflicted and the intact ROIs had shown that there are significant differences in a number of parameters. This makes it possible to develop clinically relevant measures for assessment of collaterals and prediction of treatment outcomes in further work.

Acknowledgements

This work was financially supported by the state non-commercial organisation “Russian Foundation for Basic Research”, the research project №18-29-26007.

References

- [1] V.L. Feigin, B. Norrving, G.A. Mensah, Global Burden of Stroke, *Circulation research* 120 (3) (2017), 439-448.
- [2] C. Zerna, G. Thomalla, B.C.V. Campbell, J. Rha, M.D. Hill, Current practice and future directions in the diagnosis and acute treatment of ischaemic stroke, *Lancet* 392(10154) (2018), 1247-1256.
- [3] S.A. Sheth, N. Sanossian, Q. Hao, S. Starkman, L.K. Ali, D. Kim, N.R. Gonzalez, S. Tateshima, R. Jahan, G.R. Duckwiler, J.L. Saver, F. Vinuela, and D.S. Liebeskind, Collateral flow as causative of good outcomes in endovascular stroke therapy, *Journal of neurointerventional surgery* 8 (2016), 2-7.
- [4] A.F. Frangi, W.J. Niessen, K.L. Vincken and M.A. Viergever, Multiscale vessel enhancement filtering, *Medical Image Computing and Computer-Assisted Intervention* 1496 (1998), 130-137.
- [5] A.M.M. Boers, R.S. Barros, I.G.H. Jansen, C.H. Slump, D.W.J. Dippel, A. van der Lugt, W.H. van Zwam, Y.B.W.E.M. Roos, R.J. van Oostenbrugge, C.B.L.M. Majoie, H.A. Marquering, Quantitative Collateral Grading on CT Angiography in Patients with Acute Ischemic Stroke, *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. RAMBO 2017, CMMI 2017, SWITCH 2017. Lecture Notes in Computer Science*, 10555 (2017). 176-184.
- [6] A.M.M. Boers, B.R. Sales, I.G.H. Jansen, O.A. Berkhemer, L.F.M. Beenen, B.K. Menon, D.W.J. Dippel, A. van der Lugt, W.H. van Zwam, Y.B.W.E.M. Roos, R.J. van Oostenbrugge, C.H. Slump, C.B.L.M. Majoie, H.A. Marquering, Value of Quantitative Collateral Scoring on CT Angiography in Patients with Acute Ischemic Stroke, *American journal of neuroradiology* 39(6) (2018), 1074-1082.
- [7] T. Jerman, F. Pernus, B. Likar and Z. Spiclin, Enhancement of Vascular Structures in 3D and 2D Angiographic Images, *IEEE Transactions on Medical Imaging* 35 (2016), 1-1.
- [8] J. Park, H. Kim, Radiomics as a Quantitative Imaging Biomarker: Practical Considerations and the Current Standpoint in Neuro-oncologic Studies, *Nuclear Medicine and Molecular Imaging*. 52 (2018), 99-108.
- [9] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A.G. Morganti, M. Bellomi, Radiomics: the facts and the challenges of image analysis, *European radiology experimental*. 2 (2018). doi: 10.1186/s41747-018-0068-z
- [10] W. Qiu, H. Kuang, J. Nair, Z. Assis, M. Najm, C. McDougall, B. McDougall, K. Chung, A.T. Wilson, M. Goyal, M.D. Hill, A.M. Demchuk, and B.K. Menon, Radiomics-based intracranial thrombus features on CT and CTA predict recanalization with intravenous alteplase in patients with acute ischemic stroke, *American Journal of Neuroradiology*. 40(1) (2019), 39-44.
- [11] B.B. Avants, N. Tustison, G. Song, Advanced normalization tools (ANTS). Release 1.5, 2011.
- [12] D. Dolotova, V. Donitova, I. Arhipov, F. Sharifullin, T. Zagriazkina, B. Kobrinskii, A. Gavrilov, “A Platform for Collection and Analysis of Image Data on Stroke, *Studies in Health Technology and Informatics*. 262 (2019), 308-311.
- [13] <https://pyradiomics.readthedocs.io>. Last access: January 10, 2020.
- [14] J.J.M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G.H. Beets-Tan, J.C. Fillon-Robin, S. Pieper, H.J.W.L. Aerts, Computational Radiomics System to Decode the Radiographic Phenotype, *Cancer Research* 77(21) (2017), e104-e107.

Automated Mapping of LEP Nursing Data to Nursing Minimum Data Sets

Renate RANEGGER^{a,b,1}, Werner O. HACKL^b, Inge EBERL^c, Dieter BAUMBERGER^a, Reto BÜRGIN^a, Elske AMMENWERTH^b

^a*Department of Research and Development, LEP AG, St. Gallen, Switzerland*

^b*Institute of Medical Informatics, UMIT – Private University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria*

^c*Department of Nursing Science University Hospital Munich, Germany*

Abstract. Nursing Minimum Data Sets (NMDS) intend to systematically describe nursing care. Until now NMDS have been populated with nursing data by manual data ascertainment which is inefficient. The objective of this work was to evaluate an automated mapping pipeline for transforming nursing data into an NMDS. We used LEP Nursing 3 data as source data and the Austrian and German NMDS as target formats. Based on a human expert mapping between LEP and NMDS, an automated data mapping algorithm was developed and implemented in an automatic mapping pipeline. The results show that most LEP nursing interventions can be matched to the NMDS-AT and G-NMDS and that a fully automated mapping process from LEP Nursing 3 data to NMDS-AT performs effectively and very efficiently. The shown approach can also be used to map different nursing classifications and to automatically transform point-of-care nursing data into nursing minimum data sets.

Keywords. nursing minimum data sets, data mapping, secondary use of nursing data, electronic health record, standardized nursing terminology, health informatics

1. Introduction

Nursing data must be documented in a standardized and structured way in Electronic Health Records (EHR) to allow meaningful data exchange in an eHealth environment. Structured electronic documentation also enables reuse of data from the nursing process for multiple purposes such as quality management or research ("Collect once, use many times" paradigm) [1]. *LEP Nursing 3* ("Leistungserfassung in der Pflege", i.e. nursing workload measurement) is a classification of around 500 nursing interventions. It allows structured care planning and documentation, together with nursing workload measurements [2]. At the moment, around 650 health care services in Germany, Switzerland, and Austria are using LEP in their EHR and for statistical evaluations of nursing data. LEP is semantically based on International Classification for Nursing Practice (ICNP) and is structured according to ISO-18104 [3].

Nursing Minimum Data Sets (NMDS) have been proposed to systematically describe nursing care and defined as '*systematic registration of the smallest possible number of unequivocally coded data, with respect to or for the purpose of nursing*

¹ Corresponding Author: Renate Ranegger, 9000 St. Gallen, Switzerland, renate.ranegger@lep.ch

practice, making information available to the largest possible group of users according to a broad range of information requirement' [4].

NMDS aim at supporting nursing managers, health policy decision-makers, public health experts, and nursing researchers [5]. They may lead to following benefits: access to comparable nursing care data on a local, regional, national, and international level; description of nursing care in different populations and variety of settings; availability of data for research activities; evaluation of costs and outcomes of nursing care; benchmarking of nursing quality indicators; extrapolation of trends in nursing care; and allocation of resources of hospitals [6, 7, 8].

In Austria, the NMDS-AT [8, 9, 10] and in Germany the G-NMDS [11] have been developed in recent years. However, up to now these NMDS were only theoretical constructs and weren't populated with nursing data due to two main reasons: First, as terminologies used in EHRs may be different from NMDS elements, a mapping between different terminology systems is necessary. Second, the usual way to populate NMDS by manual data ascertainment is time- and resource-intensive.

The objective of this work was to evaluate an automated mapping pipeline for transforming nursing data documented via *LEP Nursing 3* into the Austrian and German NMDS. This pipeline should comprise an automated mapping process for transforming *LEP Nursing 3* data into an NMDS representation and then loading it into the respective Nursing Minimum Data Set.

2. Methods and Materials

2.1. Human Expert Mapping between LEP and NMDS

For the manual expert mapping, three raters with LEP and NMDS expert knowledge were recruited. First, a training session was held to brief the mapping method. In the following first mapping round the source terminology, *LEP Nursing 3.3.1*, was matched to the target formats NMDS-AT and G-NMDS independently by all three raters. After that, questions which emerged during the first round and uncertainties concerning special mapping rules were discussed and solved in a further briefing session. Then, in a second mapping round, the expert mapping was finalized independently by all experts. Resulting discrepancies were clarified in a subsequent consensus process [12].

LEP nursing interventions offer sufficiently structured information for the mapping because LEP is based on ISO reference terminology. The NMDS nursing interventions provide additional structural elements compared to LEP and ISO. For example, LEP defines the intervention "Performing a full body wash" which related to the NMDS term "Body care washbasin/bed/incubator, partial assistance"; in this case, NMDS distinguishes the degree of dependency as an additional attribute. Therefore, the mapping was supplemented with data describing the patient condition using ICNP and the International Classification of Diseases (ICD). For example, nursing assessment data, nursing diagnoses or medical information were added to guide the mapping.

As the mapping of LEP to NMDS includes 1:n matches, interrater reliability was assessed using a specially defined metric. The metric for each mapped concept ranged from 0 (no match) to 1 (full match). Following scores were assigned: (a) In cases of identical mappings including the cases where the rater did not find a mapping, the agreement score was set to 1; (b) in cases of partial agreement or the case where one rater did not find a mapping but the other did, a score of 0.5 points was assigned; and (c) if

the ratings are completely distinct, 0 points were assigned. Intra- and interrater reliability were defined as the average of these scores.

2.2. Automated Data Mapping Algorithm

Based on the human expert mapping, an algorithm for an automated process for a mapping from LEP to NMDS-AT was defined [12]. NMDS-AT was chosen here as a first step, it is planned to extend the algorithm to G-NMDS later on. This algorithm comprised a five-stage mapping cascade and was implemented using Talend Open Studio (TOS, www.talend.org).

In the first stage, all the cases in which a clear 1:1 mapping between LEP and NMDS-AT was possible, were implemented (LEP→NMDS-AT). In the second stage, medical information (ICD-10 codes of the primary and all secondary diagnoses of the respective patient case) was used to make the mapping decision (LEP + ICD→NMDS-AT). For example, the NMDS category “Care of urinary stoma” will be triggered by LEP “Attending to stoma” and ICD-10 “N00-N99: Diseases of the genitourinary system”, “Z43.5 Attention to cystostomy” or “Z43.6. Attention to other artificial openings of urinary tract”. In the third stage, the nursing information of the remaining cases (all relevant ICNP codes of the patient case) was evaluated to make the mapping decision (LEP + ICNP→NMDS-AT). In the fourth stage, a decision was made if certain combinations of LEP codes, e.g. “Monitoring cuff pressure/tube” together with “Attending to the tracheal cannula” or “Attending to the tracheal catheter” were observed in a given patient case (LEP + LEP'→NMDS-AT). In the last stage, all remaining LEP elements that were not assignable to a corresponding NMDS-AT category were logged as ‘not-mappable’.

2.3. Test Data Set and Automated Mapping

To test the mapping, an anonymized data set comprising 4,390 patient cases from 7 hospital wards (from 6 different special fields) was sampled. This dataset included 789,723 documented LEP interventions (329 different instances of LEP interventions). As this dataset, due to anonymization and other organizational requirements, did not comprise other patient data such as demographic information, ICD diagnoses or ICNP data, a set of 2,000 fictitious test-patients was created. Age, ranging from 0 to 100 years, was randomly assigned and gender proportion was set to 50:50. The 4,390 real, anonymized cases were then randomly assigned to the test-patients. One primary diagnosis from the ICD-10 catalogue was randomly assigned to each case and 3,000 secondary diagnoses, as well as 3,000 ICNP codes, were randomly assigned to all cases. The test cases were stored in a PostgreSQL database for extraction and further processing.

The automated mapping procedures were compiled and exported from TOS as runnable jar files and executed on a MacBook Pro (15-inch, 2017) with 2.9 GHz Intel Core i7 CPU, 16 GB of RAM (2,333MHz LPDDR3) and macOS Mojave. The mapped cases were collected in a separate PostgreSQL database holding the NMDS-AT.

3. Results

3.1. Human Expert Mapping, Intra- & Interrater Reliability

A total of 68.8% (369/536) respectively 71.5% (383/536) instances of *LEP Nursing 3* interventions could be matched to 100% of the NMDS-AT (80/80) and G-NMDS (91/91) nursing interventions.

Table 1 shows the intra- and interrater reliability of the NMDS-AT and G-NMDS mappings. The intra-rater reliability from the NMDS-AT is somewhat higher than that of the G-NMDS (e.g. Intra-rater, Rater 3: NMDS-AT 94.0% > G-NMDS 85.5%). The reason for this may be the lower number of nursing interventions in NMDS-AT (NMDS-AT, n=80 < G-NMDS, n=91). The results of the interrater reliability from mapping round 2 suggest good agreement with 88.4% at the NMDS-AT and 88.3% at the G-NMDS.

Table 1. Average percentage of agreement (%) from the NMDS-AT and the G-NMD in consideration of partial matches (identical mapping = 1; partial mapping = 0.5; distinct mapping = 0) by 3 raters and 2 mapping rounds

	Intra-rater Rater 1	Intra-rater Rater 2	Intra-rater Rater3	Interrater Round 1	Interrater Round 2
NMDS-AT	77.0%	75.3%	94.0%	72.8%	88.4%
G-NMDS	68.0%	80.5%	85.4%	66.8%	88.3%

3.2. Automated Mapping, Outcomes & Performance

The test data set comprised 4,390 real patient cases that had been randomly assigned to 2,000 fictitious patients. The average number of cases per patient was 2.5 (\pm 1.3). Approximately 50 percent of the cases had at least one secondary ICD-10 diagnose (2,355/4,390) respective one ICNP code at least (2,215/4,390).

All *LEP Nursing 3* interventions were aggregated to instances per case and counted. The aggregation resulted in 104,278 distinct interventions for all 4,390 patient cases. Of these cases, 91 percent could be mapped (3,998/4,390) and were found in the NMDS-AT. The number of patients that were mapped to the NMDS-AT was slightly lower at 86 percent (1,729/2,000). The mapping of interventions resulted in a successful match of 47 percent of all instances (49,188/104,278). From these, 98% (48,204/49,188) were mapped in stage 1 (LEP→NMDS-AT). Stage 2 (LEP + ICD→NMDS-AT) accounted for 40 matches, 681 matches were done at stage 3 (LEP + ICNP→NMDS-AT) and stage 4 of the mapping (LEP + LEP'→NMDS-AT) yielded 84 matches.

The execution time of the whole mapping pipeline including data extraction, aggregation, mapping, loading and logging was below one second.

4. Discussion

In our study, we could show that a fully automated mapping process from *LEP Nursing 3* intervention data to elements of an NMDS-AT does work.

The manual expert mapping built a solid base for the implementation of the automated mapping algorithm. However, just under a third of the *LEP Nursing 3* interventions could not be mapped to NMDS elements by the experts. Mostly, these cases were special nursing interventions not covered by the NMDS, such as “Performing acupuncture”, indirect nursing interventions, e.g. “Carrying out surgical hand disinfection”, or nursing interventions, such as “Organising discharge”. Such nursing

interventions without assignment will be reviewed for future developments of the NMDS. The automated mapping performed best in the first stage where direct 1:1 mapping was possible. In the following mapping stages that require additional medical and nursing data the mapping success was suboptimal. This was caused by the fact that sex, age, ICD and ICNP codes had been randomly assigned to the fictitious patients. This shows that the mapping in these stages is very context-specific. With real patient cases and plausible combinations of LEP nursing interventions and medical and other nursing information, we would expect significantly improved mapping results.

5. Conclusion and Outlook

We successfully implemented and tested an automated mapping from LEP to NMDS-AT. We expect that it would be equally successful for mapping LEP to G-NMDS and comparable NMDS. Our approach is not only feasible for mapping LEP to NMDS, but also for other classification systems. All that has to be done is to define the mapping rules by human experts and then to implement the mapping cascade algorithm. Our mapping pipeline supports the transformation of nursing data from EHRs into a minimum data set representation for multiple reuse purposes on a fully automated basis.

References

- [1] Cimino JJ. Collect Once, Use Many: Enabling the Reuse of Clinical Data through Controlled Terminologies. *Journal of AHIMA*. 2007 Feb;78(2):24-9.
- [2] Oertle M, Baumgartner A. Ausmass der vollständigen Automatisierung von Pflegeleistungserfassungen basierend auf der elektronischen Patientenakte. *Swiss Medical Informatics*. 2010;26(69):37-40.
- [3] Baumberger D, Stadler M, Buchmann D. Mapping LEP Nursing 3 and ISO reference terminology model for nursing action (ISO-RTM). *Swiss Medical Informatics*. 2012;28:47-8.
- [4] Sermeus W, Delesie L, Van Landuyt J, Wuyts Y, Vanden Boer G, Manna M. The Nursing Minimum Data Set in Belgium: a basic tool for the tomorrow's health care management: Katholieke Universiteit Leuven. Centre for Health Services Research, 1994.
- [5] Sermeus W, Goossen W. Nursing Minimum Data Set. In: Mantas J, Hasman A, editors. *Textbook in health informatics: a nursing perspective*. IOS Press; 2002. p. 98-109.
- [6] Werley H, Devine E, Zorn C, Ryan P, Westra B. The Nursing Minimum Data Set: abstraction tool for standardized, comparable, essential data. *American Journal of Public Health*. 1991 Apr;81(4):421-6.
- [7] Goossen W, Epping PJ, Feuth T, Dassen TW, Hasman A, Van Den Heuvel WJ. A comparison of nursing minimal data sets. *Journal of the American Medical Informatics Association*. 1998 Mar;5(2):152-63.
- [8] Ranegger R, Ammenwerth E. Nursing Minimum Data Sets (NMDS) - eine Literaturübersicht bezüglich Zielsetzungen und Datenelemente. *Pflege* 2014 Dec;27(6):405-25.
- [9] Ranegger R, Hackl WO, Ammenwerth E. Development of the Austrian Nursing Minimum Data Set (NMDS-AT): The third Delphi-round, a quantitative online survey. *Stud Health Technol Inform*. 2015 Jan;212:73-80.
- [10] Ranegger R, Hackl WO, Ammenwerth E. Implementation of the Austrian Nursing Minimum Data Set (NMDS-AT): A Feasibility Study. *BMC Med Inform Decis Mak*. 2015 Sep;17(15):75.
- [11] Eberl I, Bartholomeyczik S. Die Übertragung des Belgischen Nursing Minimum Data Set II (B-NMDS II) auf bundesdeutsche Krankenhäuser. Ergebnisse der ersten Untersuchungsphase zum Übersetzungs- und Adaptionsprozess des Instruments. *Pflege* 2010;23(5):309-19.
- [12] Ranegger R, Eberl I, Baumberger D, Hackl WO. Automatisiertes Datenmapping zwischen Leistungserfassung in der Pflege (LEP®), dem Nursing Minimum Data Set aus Österreich (NMDS-AT) und dem Nursing Minimum Data Set aus Deutschland (G-NMDS). *Proceeding of the European Nursing Informatics (ENI)*; 2017 Sep 14-15; Hall in Tirol; Austria.

Automated Spelling Correction for Clinical Text Mining in Russian

Ksenia BALABAEVA^{a,1}, Anastasia FUNKNER^a and Sergey KOVALCHUK^a

^a*ITMO University, Saint Petersburg, Russia*

Abstract. The main goal of this paper is to develop a spell checker module for clinical text in Russian. The described approach combines string distance measure algorithms with techniques of machine learning embedding methods. Our overall precision is 0.86, lexical precision - 0.975 and error precision is 0.74. We develop spell checker as a part of medical text mining tool regarding the problems of misspelling, negation, experienter and temporality detection.

Keywords. clinical texts, electronic health records, natural language processing, spellchecker, Russian, anamnesis, misspelling correction, word embeddings

1. Introduction

Predictive modeling in medicine and healthcare is developing rapidly, and the quality of modeling results crucially depend on the quality of the input data. In these domains, a significant proportion of the information is presented with texts in natural language. Feature extraction from medical texts is challenging due to the following problems: misprints in medical terms, negation, temporality, and a person (experienter) that have experienced an event (it could be a patient or his or her relative. In the present study, our goal is to develop module of spelling correction to more accurately process and extract knowledge from clinical records in natural language.

2. Related works

Processing of clinical texts in natural language is a well-studied problem, mainly for the English language. There are many tools for labeling text, extracting entities, disease's cases, temporal and negation detection (UMLS, UIMA, IBM Watson, Apache Ruta, etc.), but their require collection of language corpus [1–4]. In some other countries, scientists develop their own corpus-free machine-learning tools or tools that can solve problems which are very specific to their

¹ Corresponding Author, ITMO University, Kronverkskyi prospect, 49, Saint Petersburg, Russia; E-mail: kyubalabaeva@gmail.com.

language [5, 6, 7]. To our knowledge, for the Russian language, a corpus of medical texts has not yet been compiled. There is a small case based on 120 records with labeled diseases and their attributes (complications, severity, treatment, etc.) [8,9]. Machine learning and artificial intelligence are increasingly used in medicine and healthcare. As a result, new methods and models are developed in NLP to obtain additional features from texts [2,4].

To extract knowledge from medical texts it is necessary to cope with many challenges: expansion of acronyms and abbreviations, interpretation of domain specific words, handling spelling mistakes, temporal tagging, negation detection [10,11], etc. There are many solutions for above mentioned problems regarding English language (UMLS, UIMA, OHNLP, etc.). However, NLP instruments for Russian are very limited and there is no generally accepted system to process clinical texts in Russian.

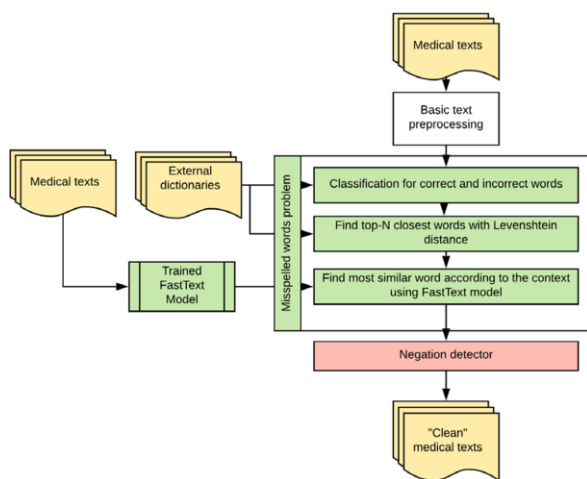


Figure 1. Modules of spelling correction and negation detection. Green blocks indicate methods described in this paper. Pink blocks are developed and implemented by colleagues.

We are developing corpus-independent modules with a spellchecker and a negation detector trained on Russian medical texts. Figure 1 shows the scheme of language processing application. And it also can be used for similar languages. In this research, we used a set of 3434 electronic health records (EHRs) of ACS patients who admitted to Almazov National Medical Research Centre (Almazov Centre) during 2010-2015. EHRs includes disease and life anamnesis, doctor's reports, discharge report, etc. in natural language. Disease anamnesis are the most unstructured records and we used them to demonstrate our approach.

In order to correct misspelled words, we need to solve three main tasks: find incorrect words in text, define the type of mistake and correct the word (Figure 1). To identify misspelled words, we use external dictionaries of russian words, medical terms, drugs, abbreviations, etc. We also broaden the pool with specific non-dictionary words of a high frequency from clinical records. Our module works with the following types of mistakes.

Table 1. Types of misspelled words.

Type of mistake	Incorrect text	Corrected text
Word reduction	anem	anemia
Missing/ misspeled symbols	anmia	anemia
Multiple words separation	hemolyticanemia	hemolytic anemia

In current realization we can automatically identify only multiple words mistake type by dividing the word sequence into all possible subsequences and evaluating if each part is an existing word and their probability to stand next to each other in a sentence is high. The basic idea of correction is to replace the incorrect word with the most similar correct word from the dict. Our approach combines machine learning technics for the word-to-vector representation with more traditional approaches of string distance calculations. To measure the performance of spell checker, we use lexical precision (which is treated as a measure of incorrect words identification), error precision and overall linguistic performance.

3. Method

The classical approaches to measure words similarity is Levenshtein distance or edit distance and its extensions (Damerau-Levenshtein-DL) [13]. It is often used for spell correction [15]. It calculates words similarity according to the minimal number of basic operations required to get the target word from the other. However, this approach doesn't consider the context provided by the sentence.

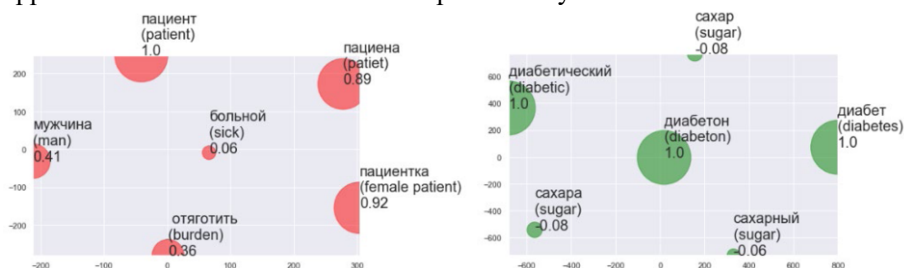


Figure 2. Most similar words for word 'patient' based on FastText vector representation (on the right). And Most similar words for the word 'diabetes' based on Word2Vec vector representation (on the left)

In order to map words into numerical vectors we use the following word-embeddings: Word2Vec [12] and also FastText [14] on SG and CBOW. We chose Word2Vec and FastText, since they are widely and efficiently used in different NLP problems, including spell correction, and leverage context of the target word. In simple words both methods build a vector representation using a hidden layer of the Neural Network. The difference between models is in token representation: FastText treats each word as a vector of n-grams, whether W2V uses the whole word. Using such representation, we can easily calculate distance

between words with cosine similarity. The example of finding close words is on the Figure 2.

4. Results and discussion

The methods were evaluated on hold-out sample of randomly sampled 200 correct medical words and 200 incorrect medical terms. All results are performed in Table 2. According to the results, we may conclude that both string distance-based algorithms and vector-based methods perform poorly when used separately. Lev. distance and DL distance are lacking context information and the ‘meaning’ of tokens. However, embedding-based approach may confuse correct word for the synonym or other related words. That is why our final model combines these approaches in the following way.

Table 2. Results of spell checker experiments

Method	Lexical Precision	Error Precision	Overall Precision
Levenshtein dist.	0.975	0.52	0.7475
Damerau–Levenshtein (DL) dist	0.975	0.545	0.76
Cos. dist. on Word2vec	0.975	0.42	0.6975
Cos. dist. on FastText CBOW	0.975	0.375	0.675
Cos. dist. on FastText SG	0.975	0.405	0.69
FastText + DL dist	0.975	0.745	0.86
Mean(CBOW, SG, W2V)+DL Dist	0.975	0.745	0.85

At the first stage, we get top-n similar words according to our word-model trained on anamnesis text, then we select m words from the dictionary and top-n words, by minimal DL distance. We also tried different ensembling approaches (voting, bagging, etc.), however it outperformed the others. The error analysis allows us to suggest, that, in general, Levenshtein distance and DL distance when used separately perform better for misprints and the words with missing letters. However, vector-based methods are better at reduced word spelling (anem, anemia). That is why it could be more precise to use independent approaches for each mistake type in future works.

5. Conclusion

We may conclude that in terms of incorrect words detection our approach works steadily and is quite precise 0.975 lexical precision. In terms of error precision it is 0.745, which is high for texts with specialized language, but still there is a room for advancement. The output data can be used for more accurate knowledge extraction from medical texts for the further processing and modelling. In order to improve our approach concerning the problem of spell checking, we are going to expand external dictionaries for more medicine names, rare medical terms,

etc. Also, we want to build ML classifier to separately work with different types of mistakes (Table 1). And our main goal in the future is to expend module's functionality towards negation, experimenter and temporality detection and make a tool for clinical text mining in Russian.

Acknowledgements

This research is financially supported by The Russian Science Foundation, Agreement #19-11-00326.

References

- [1] J. Jonnagaddala, S.T. Liaw, Coronary artery disease risk assessment from unstructured electronic health records using text mining, *J. Biomed. Inform.* **58** (2015) S203–S210. DOI:10.1016/j.jbi.2015.08.003.
- [2] N. Mehta, and M. V. Devarakonda, Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey?, *J. Allergy Clin. Immunol.* **141** (2018) 2019–2021.e1. DOI:10.1016/j.jaci.2018.02.025.
- [3] E. Ford, J.A. Carroll, Extracting information from the text of electronic medical records to improve case detection: A systematic review, *J. Am. Med. Informatics Assoc.* **23** (2016) 1007–1015. DOI:10.1093/jamia/ocv180.
- [4] C.C. Huang, and Z. Lu, Community challenges in biomedical text mining over 10 years: Success, failure and the future, *Brief. Bioinform.* **17** (2016) 132–144. DOI:10.1093/bib/bbv024.
- [5] F.R. Lucini, F. S. Fogliatto, Text mining approach to predict hospital admissions using early medical records from the emergency department, *Int. J. Med. Inform.* **100** (2017) 1–8. DOI:10.1016/j.ijmedinf.2017.01.001.
- [6] W. Sun, Z. Cai, Data processing and text mining technologies on electronic medical records: A review, *J. Healthc. Eng.* **2018** (2018). DOI:10.1155/2018/4302425.
- [7] A. Névél, H. Dalianis, Clinical Natural Language Processing in languages other than English: Opportunities and challenges, *J. Biomed. Semantics.* **9** (2018) 1–13. DOI:10.1186/s13326-018-0179-8.
- [8] A. Shelmanov, I. Smirnov, Information extraction from clinical texts in Russian, *Komp'yuternaya Lingvistika i Intellektual'nye Tehnol.* **1** (2015) 560–572.
- [9] A. Baranov, L. Namazova-Baranova, Technologies for Complex Intelligent Clinical Data Analysis, *Вестник Российской Академии Медицинских Наук.* **71** (2016) 160–171. DOI:https://doi.org/10.15690/vramn663.
- [10] M. Kreuzthaler, M.D. Bloice, A comparison of different retrieval strategies working on medical free texts, *J. Univers. Comput. Sci.* **17** (2011) 1109–1133.
- [11] D. Demner-Fushman, W.W. Chapman, and C.J. McDonald, What can natural language processing do for clinical decision support?, *J. Biomed. Inform.* **42** (2009) 760–772. DOI:10.1016/j.jbi.2009.08.007.
- [12] M. Korobov, Morphological analyzer and generator for Russian and Ukrainian languages, in: *Analysis of Images, Social networks and Texts*, 2015. DOI:10.1007/978-3-319-26123-2_31.
- [13] E. Ristard, Learning String Edit Distance, in: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, DOI: 10.1109/34.682181
- [14] P. Bojanowski, E. Grave et. al, Enriching Word Vectors with Subword Information, 2016, in [arXiv:1607.04606](https://arxiv.org/abs/1607.04606)
- [15] A. Sorokin, T.Shavrina, Automated spelling correction for russian social media texts, *Computational Linguistics and Intellectual Technologies*, 2016

Automatic Classification of Discharge Letters to Detect Adverse Drug Reactions

Vasiliki FOUFI^{a,1}, Kuntheavy ING LORENZINI^b, Jean-Philippe GOLDMAN^a,
Christophe GAUDET-BLAUVIGNAC^a, Christian LOVIS^a and Caroline SAMER^b

^a*Division of Medical Information Sciences, Geneva University Hospitals & University of Geneva, Switzerland*

^b*Clinical Pharmacology and Toxicology, Geneva University Hospitals, Switzerland*

Abstract. Adverse drug reactions (ADRs) are frequent and associated to significant morbidity, mortality and costs. Therefore, their early detection in the hospital context is vital. Automatic tools could be developed taking into account structured and textual data. In this paper, we present the methodology followed for the manual annotation and automatic classification of discharge letters from a tertiary hospital. The results show that ADRs and causal drugs are explicitly mentioned in the discharge letters and that machine learning algorithms are efficient for the automatic detection of documents containing mentions of ADRs.

Keywords. Adverse drug reaction, pharmacovigilance, text mining, document classification, supervised machine learning

1. Introduction

Adverse drug reactions (ADRs) affect 7 to 17% of hospitalized patients [1,2] and can result in serious morbidity, mortality and high costs. They are largely underreported, making active pharmacovigilance useful. The detection of ADRs can be performed through the review of electronic medical records (EMR) [3] or regular ward visits by a trained health professional [4], which can be time-consuming. In this context, data mining techniques, focusing on the automated identification of ADRs from the patient EMR can be helpful [5,6]. These techniques include structured data analysis as well as text mining, including natural language processing (NLP) [7,8]. In this context, techniques for the automatic classification of clinical documents have been proven effective [9–11].

The aim of this study is to assess the feasibility of using NLP techniques to detect the presence or absence of ADRs in discharge letters written in French and extracted from patients hospitalized in a tertiary hospital via a hybrid –machine learning and rule-based– method. In this paper, we will present the supervised learning method. Particularly, three machine learning algorithms for document classification have been applied and evaluated. For the creation of the training and test datasets, manual processing of 300 discharge letters was performed. The results show that ADRs are reported in the documents and that NLP tools are efficient for their automatic detection.

¹ Corresponding Author, Division of Medical Information Sciences, Geneva University Hospitals & University of Geneva, Switzerland; E-mail: vasiliki.foufi@unige.ch.

2. Method

2.1. Data collection

Our study was approved by the local ethics committee (study number: 2016-02107). Hospitalized adults for whom a specialized consultation from clinical pharmacologists (in 2015 and 2016) had identified the occurrence of serious ADRs, during or leading to hospitalisation, were included in the study. Out-patients and cases of non-serious ADRs were not included. Based on these criteria, a dataset of 100 positive discharge letters (presence of ADR) and 200 negative letters (absence of ADR) was constituted.

2.2. Data processing

2.2.1. Manual annotation

For the creation of the training and test datasets, an expert manually annotated 100 discharge letters (positive dataset) and validated 200 letters (negative dataset). Based on specific guidelines, sequences of the following categories were annotated:

1. Drugs

The drugs category is divided in 3 sub-categories: a) commercial names, b) international nonproprietary names (INN), c) therapeutic class.

2. ADRs

Occurrences of ADRs and their consequences, symptoms, laboratory values are annotated. ADRs are divided in 3 sub-categories: a) names (*hépatite/hepatitis*), b) periphrases (*perturbation des tests hépatiques/liver test abnormalities*), c) characteristics (*hémoglobine à 75 g/l/haemoglobin at 75 g/l*).

3. Trigger words

Words like *imputabilité/causality*, *stoppé/stopped*, *suspect/suspect* that imply the presence of an ADR are annotated.

4. Drug indications

Indications for drugs entailed in ADRs are annotated.

2.2.2. Automatic classification of discharge letters

For the automatic classification of the discharge letters into positive or negative, a supervised learning approach was followed. The dataset is composed of the positive letters containing at least one annotation of ADR and the negative letters validated by the expert. For this task, three machine learning algorithms widely used for text classification tasks were applied: Support Vector Machine (SVM), Naïve Bayes Classifier, and Linear Classifier. From the whole dataset, 80% was used for training and 20% for testing.

3. Results

3.1. Manual annotation

Out of 100 letters of the positive dataset, 87 letters contained at least one annotated sequence. The mean length of a discharge letter is 785 words. In total, 1471 sequences were annotated. These results are summarized in Table 1:

Table 1. Number of occurrences per annotation category.

Annotation category	Number of occurrences	Unique occurrences
Commercial name	170	76
International nonproprietary name	210	87
Therapeutic class	126	59
Trigger word	293	156
ADR	441	217
Characteristic	130	103
Drug indication	37	28
Periphrasis	64	35

From the 200 discharge letters considered as negative (absence of ADR), 47 reported an ADR and 2 were empty. Therefore, the final negative dataset consists of 151 letters.

3.2. Automatic document classification

Three well-known classification methods implemented in the Scikit-learn python library [12] were applied and compared: Support Vector Machine, Naive-Bayes, and Linear Classifier. For the SVM model, the radial basis function was selected with a ‘scale’ gamma. The Multinomial Naive-Bayes was trained with default parameters. Eventually, the Linear Classifier is based on a stochastic gradient descent with a lower stopping criterion than default (tol=1e-6). Figure 1 represents the classification accuracy of each model after 50 iterations. For each iteration, the test set represents 20% of the whole corpus, i.e. 17 positive documents and 30 negative documents (k-fold = 0.8). The training time is ~100ms for each SVM iteration whereas it takes 2mn for both the Naïve Bayes and the Linear Classifier.

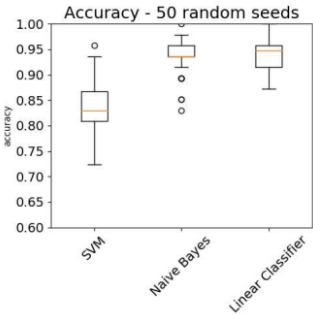


Figure 1. Mean accuracy over 50 iterations.

The SVM classifier achieved 0.83 accuracy, Naïve Bayes achieved 0.94 accuracy and the Linear Classifier 0.94. Complete results are shown in Table 2:

Table 2. Evaluation results of the automatic classification.

Classification method	Class	Precision	Recall	F1 score	Test set
SVM	Positive	0.80	0.97	0.88	17
	Negative	0.92	0.60	0.73	30
Naïves Bayes	Positive	0.93	0.97	0.95	17
	Negative	0.95	0.88	0.91	30
Linear Classifier	Positive	0.97	0.93	0.95	17
	Negative	0.90	0.95	0.92	30

The confusion matrices in Figures 2-4 display the performance of each classifier at the classification task:

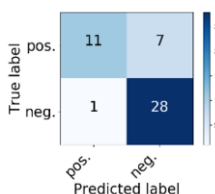


Figure 2. SVM.

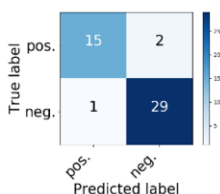


Figure 3. Naïve Bayes.

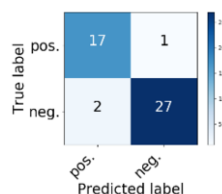


Figure 4. Linear Classifier.

4. Discussion

The manual annotation of positive discharge letters by an expert showed that ADRs were explicitly mentioned in most cases (>80%). A significant bias is that the test population were patients who had already received a specialized pharmacology consultation that had identified the ADR, thereafter mentioned in the discharge letter. Drugs were almost equally mentioned as commercial names, INN and therapeutic classes, and this has to be taken into account for their automated detection, especially given the diversity of commercial names in different countries. Trigger words were frequently present (293 occurrences); therefore, they constitute useful tools for the automatic detection of ADRs. ADRs were most frequently mentioned as plain terms, such as MedDRA (Medical Dictionary for Regulatory Activities) derived terms (i.e. *hepatitis*), but were also described as periphrases or as laboratory characteristics in many cases (approximately 200 occurrences) which makes their automatic detection challenging given the fact that the distinction between the 3 sub-categories was not always straightforward even for the human annotator.

For the automatic classification task into positive and negative discharge letters, three machine learning algorithms were applied and evaluated on the dataset. Naïve Bayes and Linear Classifier achieved the same mean accuracy over 50 iterations (0.94) and high precision and recall (Table 2).

A major limitation of this study is that the dataset was manually processed by only one annotator. Also, the classifiers should be applied and evaluated on a larger dataset.

5. Conclusion

In this study, we presented the methodology used for the manual and automatic processing of discharge letters generated in a tertiary hospital in the aim to automatically detect the presence or absence of ADRs. A dataset of 300 discharge letters written in French was manually processed and the output was used to train and test three machine learning algorithms for document classification. After comparison, we concluded that Naïve Bayes and Linear Classifier performed better than SVM at this task.

The manual annotation of the dataset from another expert will serve to create a gold standard corpus. In a next step, the trigger words and sequences describing the presence of ADRs identified during the manual annotation will be included in the rules that are being developed for the automatic identification and extraction of ADRs and their relations with causal drugs. Then, the hybrid method –machine learning and rule-based– will be applied and evaluated on the gold standard dataset.

Acknowledgements

This research is funded by the HUG's Private Foundation and the HUG's Medical Director PRD (Projets Recherche & Développement) project funding.

References

- [1] Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. - PubMed - NCBI, (n.d.). <https://www.ncbi.nlm.nih.gov/pubmed/9555760> (accessed October 15, 2019).
- [2] Frequency of adverse drug reactions in hospitalized patients: a systematic review and meta-analysis. - PubMed - NCBI, (n.d.). <https://www.ncbi.nlm.nih.gov/pubmed/22761169> (accessed October 15, 2019).
- [3] Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register, (n.d.). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4537312/> (accessed October 15, 2019).
- [4] Methods and systems to detect adverse drug reactions in hospitals. - PubMed - NCBI, (n.d.). <https://www.ncbi.nlm.nih.gov/pubmed/11735652> (accessed October 15, 2019).
- [5] Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? - PubMed - NCBI, (n.d.). <https://www.ncbi.nlm.nih.gov/pubmed/19757412> (accessed October 15, 2019).
- [6] Data-mining-based detection of adverse drug events. - PubMed - NCBI, (n.d.). <https://www.ncbi.nlm.nih.gov/pubmed/19745372> (accessed October 15, 2019).
- [7] Using text-mining techniques in electronic patient records to identify ADRs from medicine use. - PubMed - NCBI, (n.d.). <https://www.ncbi.nlm.nih.gov/pubmed/22122057> (accessed October 15, 2019).
- [8] Automated detection of adverse events using natural language processing of discharge summaries. - PubMed - NCBI, (n.d.). <https://www.ncbi.nlm.nih.gov/pubmed/15802475> (accessed October 15, 2019).
- [9] Clinical Document Classification Using Labeled and Unlabeled Data Across Hospitals. - PubMed - NCBI, (n.d.). <https://www.ncbi.nlm.nih.gov/pubmed/30815095> (accessed October 15, 2019).
- [10] Machine learning in automated text categorization, (n.d.). <https://dl.acm.org/citation.cfm?id=505283> (accessed October 15, 2019).
- [11] K.J. Dreyer, M.K. Kalra, M.M. Maher, A.M. Hurier, B.A. Asfaw, T. Schultz, E.F. Halpern, and J.H. Thrall, Application of recently developed computer algorithm for automatic classification of unstructured radiology reports: validation study, *Radiology*. **234** (2005) 323–329. doi:10.1148/radiol.2341040049.
- [12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*. **12** (2011) 2825–2830.

Automatic Extraction of Risk Factors for Dialysis Patients from Clinical Notes Using Natural Language Processing Techniques

George MICHALOPOULOS^a, Hammad QAZI^a, Alexander WONG^a, Zahid BUTT^a, and Helen CHEN^{a,†}

^a University of Waterloo

Abstract. Studies have shown that mental health and comorbidities such as dementia, diabetes and cardiovascular diseases are risk factors for dialysis patients. Extracting accurate and timely information associated with these risk factors in the patient health records is not only important for dialysis patient management, but also for real-world evidence generation. We presented HERALD, an natural language processing (NLP) system for extracting information related to risk factors of dialysis patients from free-text progress notes in an electronic dialysis patient management system. By converting semi-structured notes into complete sentences before feeding them into the NLP module, the HERALD system was able achieved 99%, 83% and 80% accuracy in identifying dementia, diabetes and infarction, respectively.

Keywords. Natural language processing, Text data mining, Real-world data, Dialysis patient risk factor

1. Introduction

The amount and complexity of data amassed in electronic health records (EHR) are enormous. EHR data are not only a vital part of providing patient care but a rich source for a wide range of secondary uses, including quality control, real world evidence generation and clinical research. Although great progress in the standardization of medical terminology has been made [1], semi-structured or unstructured text data still constitute a large portion of EHR data. Thus, one of the challenges facing real-world evidence generation is the automatic extraction of high quality information from narrative data. However, rapid advances in the field of Natural Language Processing (NLP) could offer viable solutions in extracting information from complex clinical text. NLP toolkits such as cTakes

[2] was capable of recognizing Unified Medical Language System (UMLS) concepts in clinical texts. NLP algorithm was shown to be effective in identifying breast cancer recurrence from clinical notes with above 90% accuracy [3]. Furthermore, Berrios et al.

[4] used a method that involved tokenization, syntactic parsing and noun phrase identification to identify UMLS concepts associated with infectious diseases.

These studies demonstrated the feasibility of using NLP techniques to significantly reduce the work of manual annotation of clinical data. Therefore, the ability to identify

[†] Corresponding Author, Helen Chen, University of Waterloo, School of Public Health and Health Systems; E-mail: helen.chen@uwaterloo.ca

key information such as events of interests and risk factors in the vast amount of narrative EHR data will be increasingly important in personalizing treatment or decision support systems [5].

HERALD is an NLP-based data extraction system, collaboratively developed by a team of clinicians at the Grand River Hospital (Kitchener, Ontario) and researchers of the University of Waterloo.

The HERALD system used the clinical notes of patients in the EHR system to identify relevant information associated with their potential risk associated with withdrawal from the dialysis treatment. More specifically, the EHR data were used to determine whether a patient’s progress notes contain documentation related to dementia, diabetes and infarction [6]. In this paper, we present the design and the validation results of the HERALD system. We demonstrated the feasibility of the HERALD system to process non-standardized narratives of progress notes, as well as semi-structured text ,i.e. templates embedded in the free-text progress notes.

This study provided insights on the utilities and limitations of the open-source NLP toolkit cTakes in processing dialysis content, as well as strategies of adapting the HERALD system to a specific clinical domain.

The study used clinical notes extracted from the renal patient management information system, Nephrocare® from 1st January 2012 to 30th September 2017 in a large tertiary care hospital in Ontario, Canada. The hospital provides a full spectrum of Chronic Kidney Disease services, including nephrology clinics, pre-dialysis, hemodialysis, peritoneal dialysis, home hemodialysis, and pre-transplantation for eligible patients living in Waterloo Region and Wellington County. The Tri-Hospital Research Ethics Board (THREB) and the University Human Research Ethics Committee approved the research project (THREB File # 2016-0619) along with the waiver for “the requirement to obtain patient informed consent”. The notes can be divided into two categories:

(i) **free-text notes**: free text narratives generated by multi-disciplinary care team members (ii) **semi-structured data**: data in a predefined template (e.g., “History of falling; within 3 months [x] NO,[JYES”). Full chart review were performed by three medically trained personnel to extract risk factors associated with dialysis withdrawal. Our study was focused on three specific risk factors: dementia, diabetes and infarction. The statistics for the positive and negative classes for each risk factor can be found in Table 1.

Table 1. statistics of the dataset

risk factor	positive class	negative class
Dementia	39	644
Diabetes	447	236
Infarction	195	488

2. Method

The HERALD system consists of three main modules: **prepossessing** module; **NLP** module and **Rule** module. The process pipeline is depicted in Fig 1.

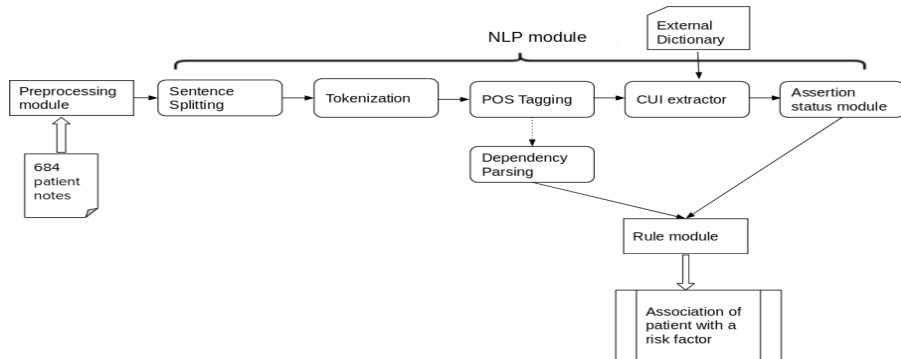


Figure 1. Pipeline of the HERALD system

The function of the **preprocessing** module is to remove the information that was not tagged as correct due to the semi-structured format of the notes. We tracked the number of repetitions of the sentences between the patients and only removed the sentences that were not connected with a filled box and can be observed for multiple patients. The logic behind this policy is that, in structured data, the sentences need to be repeated between the patient forms and cannot change over multiple patients. Therefore, unique sentences are intentionally added by a physician for a specific patient.

For the extraction of information from the unstructured notes, we developed an **NLP** module with the open-source Apache clinical Text Analysis and Knowledge Extraction System (cTakes) [2]. cTakes is a natural language processing tool, whose components are specifically trained for the clinical domain. Our NLP pipeline was composed of (i) A *concept coding* module where we tried to identify terms of interest for which a concept unique identifier (CUI) exists in UMLS [1]. In this module, the sentences were split into tokens and the tokens were normalized to their based form. A POS tagging sub-module was then used to tag the normalized tokens to their part-of-speech and finally a named-entity recognition algorithm used this information to identify terms of interest that exist in the UMLS library. (ii) An *assertion status* annotation module where additional information for each CUI were discovered in the texts, such as whether a term is negated or whether it expresses uncertainty (iii) A *coreference annotation* [7] module which provides information about whether the person of interest for a specific CUI is the patient or a family member or someone irrelevant. Finally, after reviewing the initial results with only the terms from the UMLS dictionary, our expert examined the notes of patients that were flagged as false-negative and he provided new terms that are indicators of a specific risk factor. We created a new custom dictionary with the additional CUI that were determined to be relevant to the identification of different risk factors.

In the **Rule** module, different rules were created in order to determine the existence of a connection between a patient and a specific risk factor. Our first rule was that a patient is diagnosed with a specific risk factor only if all the conditions below were true:

(i) A term/s that was/were connected to this risk factor was found in his notes (ii) The term/s was/were not negated (iii) The term/s did not express uncertainty (iv) The subject of the term/s was/were the patient. Finally, we investigated if a term was connected to the following sentences by using the dependency parser, which was provided by cTakes[8]:

(i) can cause, (ii) risk of, (iii) voicing concerns about, (iv) raising concerns about, (v) no prior history of, and (vi) free. We chose these sentences as they express uncertainty or negation but cTakes failed to identify them. By analyzing the structure of the dependency

tree, our system could understand the terms that were connected to the above sentences and if they were indeed connected to one of them, then we did not consider these terms as proof that a patient is connected to a specific risk factor. Python 3.7 was used for the creation of this module and the preprocessing module.

3. Results

Cardiac disease [9] and dementia [10] were found to be associated with increased risk of death and dialysis withdrawal. Based on the quality of source data, number of patients for each risk factor and the definition of each risk factor, we tested how accurately the HERALD system could determine the connection of dialysis patients with dementia, diabetes, infarction in the progress notes. HERALD processed the notes of each patient in order to identify each risk factor and the results were compared with the manual chart review in order to calculate the classification accuracy. The accuracy of our system was reported in Table 2.

Table 2. Results of experiments

risk factor	Accuracy	Sensitivity	Specificity
Dementia	0.9941	0.8333	1.000
Diabetes	0.8375	0.8318	0.8481
Infarction	0.8038	0.7077	0.8422

4. Discussion

The HERALD system identified a patient with dementia with high sensitivity and good specificity. Given dementia was considered as a contraindication for dialysis treatment, dementia was not as common as other risk factors such as diabetes and infraction in the dialysis population. The HERALD system was able to identify accurately the presence of of dementia in a clinical note, which could help care providers to provide needed support for patients or their families.

One of the limitations of the HERALD system, or any NLP systems in general, is the ability to create generalizable rules that could capture the experience of a medical expert. For example, one false-negative case in our experiment is due to the more complex reasons associated with the inference of the presence of dementia, such as: (i) The patient had delirium (ii) No reversible causes of cognitive and motor decline (iii) Overall history suggestive of dementia. Furthermore, it should be noted that during his chart review, in cases of uncertainty, our medical expert also consulted other clinical reports available in a regional EHR system. These cases were kept as a reminder that, in a real-world scenario, relevant information can be absent from clinical notes and a decision support system should take into consideration multiple sources - structured and unstructured data.

5. Conclusion

Our study demonstrated the feasibility of using an NLP-based system to identify the association of a patient with specific risk factors. This study also proved that the NLP

technique could be a useful tool for identifying passages containing relevant information of interests. However, we also showed that an NLP system could not fully replace the judgment of a clinical expert. Since it is possible to automatically extract structured data from the dialysis program's EHR system (e.g. NephroCare), combining structured data and features extracted from notes, could be a promising future direction to improve the performance of the HERALD system. Also, we aim to combine all available data extracted from the EHR system, including structured and non-structured data in order to build machine learning models to predict if a patient is likely to withdraw from a dialysis treatment. Finally, our future plan is to investigate different feature selection techniques as it is a crucial process for the performance of machine learning models [11].

Acknowledgments

We thank Tina Mah, Kathleen Lavoie, Peter Varga, and Kim Hendrick at the Grand River Hospital for enabling access to data and providing insights on dialysis care. We are grateful for the support of the MITACS and Roche Canada for providing Postdoc Fellowship to Hammad Qazi.

References

- [1] Spackman KA, Campbell KE, Co'te' RA. SNOMED RT: a reference terminology for health care. Proc AMIA Annu Fall Symp. 1997;640-4. PMID: 9357704; PMCID: PMC2233423.
- [2] Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association. 2010 Sep 1;17(5):507-13.
- [3] Carrell DS, Halgrim S, Tran DT, Buist DS, Chubak J, Chapman WW, Savova G. Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence. American journal of epidemiology. 2014 Mar 15;179(6):749-58
- [5] Berrios DC, Kehler A, Fagan LM. Knowledge requirements for automated inference of medical textbook markup. Proc AMIA Symp. 1999;676-80. PMID: 10566445; PMCID: PMC2232726.
- [7] Hornberger J. Electronic health records: a guide for clinicians and administrators. JAMA. 2009 Jan 7;301(1):110-.
- [8] Qazi HA, Chen H, Zhu M. Factors influencing dialysis withdrawal: a scoping review. BMC Nephrol. 2018 Apr 24;19(1):96. doi: 10.1186/s12882-018-0894-5. PMID: 29699499; PMCID: PMC5921369.
- [9] Zheng J, Chapman WW, Miller TA, Lin C, Crowley RS, Savova GK. A system for coreference resolution for the clinical narrative. J Am Med Inform Assoc. 2012 Jul-Aug;19(4):660-7. doi: 10.1136/amiainl-2011-000599. Epub 2012 Jan 31. PMID: 22298565; PMCID: PMC3384116.
- [11] Choi JD, Nicolov N. K-best, locally pruned, transition-based dependency parsing using robust risk minimization. Recent Advances in Natural Language Processing V. 2009 Oct 22;309:205-16.
- [12] Wetmore JB, Yan H, Hu Y, Gilbertson DT, Liu J. Factors associated with withdrawal from maintenance dialysis: a case-control analysis. American Journal of Kidney Diseases. 2018 Jun 1;71(6):831-41.
- [13] Kurella M, Mapes DL, Port FK, Chertow GM. Correlates and outcomes of dementia among dialysis patients: the Dialysis Outcomes and Practice Patterns Study. Nephrology Dialysis Transplantation. 2006 Sep 1;21(9):2543-8.
- [14] Bates J, Fodeh SJ, Brandt CA, Womack JA. Classification of radiology reports for falls in an HIV study cohort. J Am Med Inform Assoc. 2016 Apr;23(e1):e113-7. doi: 10.1093/jamia/ocv155. Epub 2015 Nov 13. PMID: 26567329; PMCID: PMC4954638..

Automating Quality Control for Structured Standardized Radiology Reports Using Text Analysis

Anjani DHRANGADHARIYA^{a,1}, Sandy MILLIUS^a, Cyril THOULY^b, Benoit RIZK^b,
Dominique FOURNIER^b, Henning MÜLLER^{a,c}, and Hugues BRAT^b

^aUniversity of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

^bInstitut de Radiologie de Sion (IRS), Groupe 3R, Sion, Switzerland

^cUniversity of Geneva, Geneva, Switzerland

Abstract. Radiology reports describe the findings of a radiologist in an imaging examination, produced for another clinician in order to answer to a clinical indication. Sometimes, the report does not fully answer the question asked, despite guidelines for the radiologist. In this article, a system that controls the quality of reports automatically is described. It notably maps the free text onto MeSH terms and checks if the anatomy and disease terms match in the indication and conclusion of a report. The agreement between manual checks of experienced radiologists and the system is high with automatic checks requiring only a fraction of time. Being able to quality control all reports has the potential to improve report quality and thus limit misunderstandings, loosing time for requesting more information and possibly avoid medical mistakes.

Keywords. Radiology Reports, Quality Control, Text Analysis

1. Background

Radiologists routinely document, report and communicate diagnostic information to the referring clinicians through structured standardized reports (SSRs). Different from unstructured documents, a typical radiology report is divided into the following sections: type of examination, clinical indication, technique, findings and conclusion [1,2]. A referring clinician conveys any identified clinical question requesting a radiology examination. This clinical question is repeated in the *indication* section of a report and the aim of a radiologist is to clearly communicate the answer to this clinical question through the *conclusion* section [3]. Despite guidelines for structuring reports, several problems can prevent clear communication between a radiologist and a clinician. One of the most frequent problems found in reports is a difference between the content of the indication and the conclusion sections whereby a report either fails to answer the clinical question of the referring clinician or has an unclear answer [4]. For example, we identified a report where a clinician requested a MRI (Magnetic Resonance Imaging) scan upon indications of tendinopathy and bursitis

¹ Corresponding Author: E-mail: anjani.dhrangadhariya@hevs.ch.

(comorbidities). The *conclusion* section confirmed tendinopathy but failed to validate or contradict the clinicians’ question about bursitis.

Such human errors could be avoided through automated quality control (Q.C.) of reports. The work of this article describes a Java-based software prototype that employs data integration and semantic comparison to identify the reports with potential differences in reporting and informs the concerned radiologist by flagging them. We assess prototype performance by measuring overall concordance between the indication and conclusion sections using a data set of radiology reports by comparing the results of automated evaluation with human evaluation.

2. Methods

2.1 Data set used

A set of 200 randomly chosen, anonymized, French-language reports was obtained from Institut de Radiologie de Sion (IRS), a private radiology center located in the French speaking part of Switzerland. An individual report comprises the following sections: indications, description and conclusions (see Figure 1).

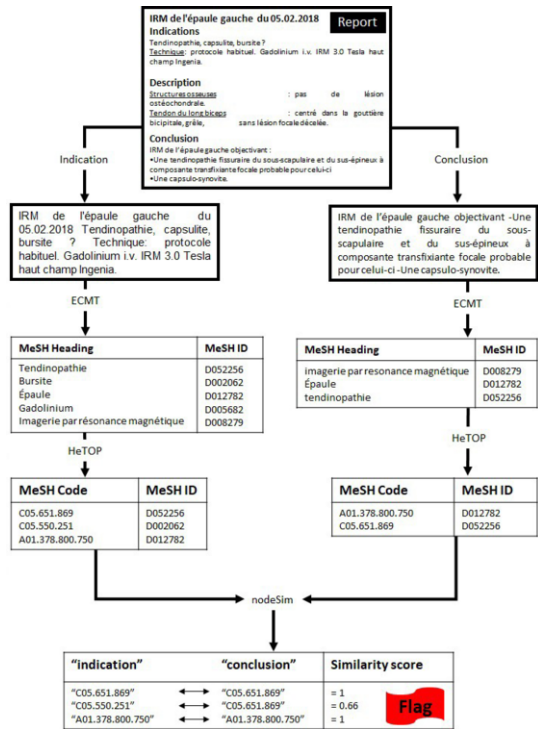


Figure 1. Picture showing the processes followed in the developed prototype.

The data set was not only used to assist prototype development but was also used to evaluate the prototype.

2.2 Medical Subject Headings

Medical Subject Headings (MeSH) are a hierarchically-organized terminology mainly used for organization and cataloguing of the biomedical information and literature. It is organized into a tree structure with its root node branching into 16 broad thematic categories like “Anatomy”, “Diseases”, “Chemical and Drugs”, etc. These categories are further divided into subcategories or MeSH headings. In this way, MeSH enforces uniformity and consistency across the terminology in a way that articles corresponding to a particular topic are indexed under a particular MeSH heading. For example., all the studies about *benign cancer* are indexed under MeSH heading *Neoplasms*². Each MeSH heading in the MeSH tree has a code indicating its location in the tree and a unique MeSH ID [5].

2.3 MeSHSim

MeSHSim is an R-package with “nodeSim” functionality that measures semantic similarity between any two MeSH codes as a function of their length and their position in the MeSH tree. The similarity values range between 0 and 1 [6].

2.4 HeTOP

HeTOP (Health Terminology/Ontology Portal)³ is a portal, housing 70 health terminologies in 32 languages including MeSH and Radiology Lexicon (RadLex). It comprises a semantically interoperable network of 2 million cross-lingual, health-domain related concepts. For example, querying the French term *radiologie* leads to its MeSH counterpart “Radiology” from other multi-lingual terminologies⁴. This rich multilingual content is available for research and can be programmatically accessed via authenticated Representational State Transfer (REST) or Simple Object Access Protocol (SOAP) requests [7].

2.5 ECMT

L'Extracteur de Concepts Multi-Terminologique (ECMT) combines a rule-based and an NLP (Natural Language Processing)-based approach to extract health-related concepts from French-language texts using French-language terminologies in HeTOP. ECMT service can be programmatically accessed via authenticated REST or SOAP requests [7,8].

3. Results

3.1 Prototype developed

The developed prototype fetches a non-empty, radiology report file from internal storage followed by rule-based extraction of free text from the indication and

² <https://meshb.nlm.nih.gov/record/ui?ui=D009369>

³ <https://www.hetop.eu/hetop/fr/?q=&home>

⁴ <https://www.hetop.eu/hetop/#rr=MSH D 011871&oti=all&q=Radiology>

conclusion sections (see Figure 1). An authenticated REST request to ECMT then extracts French-language health-related concepts from the free text. ECMT returns an XML (eXtensible Markup Language) file that is parsed to extract only the MeSH Headings and unique ID's corresponding to Anatomy [A] and Disease [C]. Then, an authenticated REST request to HeTOP returns an XML file with MeSH codes and paths corresponding for the requested MeSH headings. Next, weighted-link semantic similarity between individual anatomy and disease codes is calculated from indication and conclusion using the nodeSim functionality from MeSHSim. If every anatomy and disease code in indication has a similarity score of 0.85 or more with their counterparts in conclusion, the report is considered as

Q.C. passed, otherwise it is flagged. A threshold of 0.85 was chosen, as it pertains to path length three or less between MeSH codes being compared, based on experiments with a few validation cases. This threshold accounts for the conceptual similarity between the MeSH headings being compared. [6]. For example, a similarity score between MeSH anatomy terms “knee” and “toes” is 0.815, but between “toes” and “hallux” it is 0.948.

3.2. Prototype evaluation

To evaluate the performance of the prototype, overall percentage concordance between indication and conclusion for all the radiology reports in the data set was automatically assessed. Out of the 200 reports in the data set, a small subset were flagged by the prototype because the text corresponding to the *conclusion* section was missing in these reports. The *conclusion* text from these reports was deliberately removed for testing this scenario. Of the remaining reports, MeSH terms and codes for the anatomy and diseases were extracted from the indication and conclusion free-texts using ECMT and HeTOP. Only 93 and 143 reports respectively contained MeSH terms corresponding to anatomy and disease. Of these 143 reports, semantic similarity between *indication* and *conclusion* MeSH terms was obtained using nodeSim and total concordance was calculated. To establish a ground truth, two expert radiologists repeated the above procedure manually and calculated the total concordance.

Upon total concordance assessment between the indication and conclusion, the prototype reported 45% concordance for the anatomy information, while manual evaluation reported a 61% concordance. For disease information, the prototype reported 80% concordance versus 88% with the peer review. Based on ground truth, the algorithm accuracy for anatomy was 84% and for disease 92%. It took the experience radiologists about 1.5 hours to conduct manual quality checks, but the prototype did it in a matter of seconds.

4. Discussion and Conclusion

Reports with perfect match for the anatomy and disease MeSH terms between indication and conclusion correctly passed Q.C, multiple reports were flagged when a particular anatomy term (e.g. knee) was mentioned in *indication*, but was missing in the *conclusion*. The prototype success, however, is subject to ability of ECMT to accurately extract the concepts from free texts and also upon availability of a particular concept in the MeSH vocabulary. Currently, the prototype accounts only for the MeSH terminology, which lacks some radiology-specific terms [9]. For example, a report

mentioned *epicondyle* in *indications* section, but not in the *conclusion* section. This report passed Q.C. because no MeSH term corresponding to *epicondyle* was identified from indications. Indeed, this term was absent in MeSH, but present in RadLex⁵ and hence complementing the two can likely improve the results for the missed radiology terms. However, only a part of RadLex is currently available in French. In another instance, a disease term "Déchirure" from free text indication was not identified by the ECMT, but a standardized, English counterpart for "Déchirure" existed in MeSH as "Rupture". This instance was reported to the ECMT team and the term was later added and semantically linked to its MeSH counterpart. Overall prototype performance should also be improved with lexical standardization of disease terms in radiology reports, and further developments in ECMT algorithm.

In conclusion, the automated prototype showed good performance compared to the expert radiologists for assessment of indication and conclusion concordance in the radiology reports. It also reduces time and cost required for the quality control. The prototype additionally identified the lack of systematic reporting of relevant anatomy information in the conclusion sections, a fact that can be highlighted in guidelines for radiologists.

References

- [1] R. B. Gunderman and L. R. McNeive, "Is structured reporting the answer?," *Radiology* **273**(1), pp. 7–9, 2014. PMID:25247561.
- [2] A. P. Brady, "Radiology reporting-from hemingway to hal?," *Insights into Imaging* **9**(2), pp. 237–246, 2018.
- [3] A. Wallis and P. McCoubrie, "The radiology report—are we getting the message across?," *Clin. Radiol.* **66**, pp. 1015–1022, Nov. 2011.
- [4] T. Laswad, D. Fournier, J. Moreau, H. Brat, M. Deac, T. Tancredi, F. Bonvin, T. Zand, and P. Benedict, "Structured reports: 10 years experience in private radiology clinics," in *European Congress of Radiology 2014*, Springer International Publishing, (Vienna, Austria), 2014.
- [5] F. B. Rogers, "Medical subject headings," *Bull Med Libr Assoc* **51**, pp. 114–116, Jan 1963.
- [6] J. Zhou, Y. Shui, S. Peng, X. Li, H. Mamitsuka, and S. Zhu, "MeSHSim: An R/Bioconductor package for measuring semantic similarity over MeSH headings and MEDLINE documents," *J Bioinform Comput Biol* **13**, p. 1542002, Dec 2015.
- [7] A. Benis, M. Crisan-Vida, L. Stoicu-Tivadar, and S. Darmoni, "A Multi-Lingual dictionary for health informatics as an international cooperation pillar," *Stud. Health Technol. Inform.* **262**, pp. 31–34, 2019.
- [8] M. H. Metzger, Q. Gicquel, D. Rabarijaona, F. Segond, A. Bittar, L. Dini, L. F. Soualmia, and S. J. Darmoni, "A secondary use of the electronic health record for epidemiological studies and evaluation of health care quality: the SYNODOS project," *Journes Francophones d'Imagerie mdicale* **1379**, January 2014.
- [9] D. Demner-Fushman, S. E. Shooshan, L. Rodriguez, S. Antani, and G. R. Thoma, "Annotation of chest radiology reports for indexing and retrieval," in *Multimodal Retrieval in the Medical Domain*, pp. 99–111, Springer International Publishing, 2015.

⁵ <http://www.radlex.org/RID/RID38647>

Bank of Digital Filters for Identification of Combined Drug Products in a French Public Database

Adrien UGON^{a,1}, Mélanie COURTINE^b, Karima SEDKI^b, Pierre QUENTIN^{a,b},
Marion VALETTE^{a,b}, and Rosy TSOPRA^{c,d,e}

^a ESIEE-Paris, Cité Descartes BP99, Noisy-le-Grand, 93162, France.

^b Inserm, Sorbonne Université, Univ Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, F-75011 Paris, France.

^c INSERM, Centre de Recherche des Cordeliers, UMRS 1138, Université Paris-Descartes, Université Sorbonne Paris Cité, France.

^d Department of Medical Informatics, Hôpital Européen Georges Pompidou, AP-HP, Paris, France

^e LTSI Université Rennes 1 Team Health Big Data, Rennes

Abstract. Drugs information systems, prescription support softwares, and drug decision support systems need to reason on drug properties. Combined pharmaceutical products need to be considered specifically because they may require a specific processing. Hence, they also need to be identified to automate the population of databases with up-to-date property values. We defined a set of digital filters designed for the identification of antibiotics in a public database. Four different filters are proposed, to be combined to extract the relevant information. Evaluation was conducted to combine filters and retrieve information about random combined antibiotics with success. However, information provided in the structured files of the French drug database is limited; information provided in the HTML files suffers from a lack of quality. Hence, reuse of this data and this information should be performed very cautiously.

Keywords. Drug Combinations, Information Retrieval, Drug databases, Antibiotics

1. Introduction

World Health Organization defines combined drugs as “*pharmaceutical products containing two or more active ingredients (incl. combination packages). Stereoisomeric mixtures are regarded as plain products. Medicinal products which in addition to one active ingredient contain auxiliary substances intended to increase the stability of the product (e.g. vaccines containing small amounts of antibacterials), increase the duration (e.g. depot formulations) and/or increase the absorption (e.g. different solvents in various dermatologicals) are considered as plain products*”. It is assumed that “the classification of combination products is a challenge in any classification system”. For instance, any support decision tool processing dose information should be able to have a

¹ Corresponding Author, E-mail: adrien.ugon@esiee.fr.

specific processing of combined products, and, thus, should be able to distinguish plain to combined medication. These automated tools are used to generate alerts and/or decision or recommendation [1].

Public databases provide information about drugs. French public Drug Database [2] provides information for each drug on the market either in structured Excel files, or as HTML pages including the Summary of Product Characteristics provided by the national drug agency and the medication package insert, provided by the pharmaceutical industry, which markets the medication. These HTML pages can be processed as text documents.

None of the fields of the structured files can be used to decide whether a medication is a combination or not. The world health organization has provided guidelines for combination products, but there is no standard rule that may be used on the basis of the ATC code to identify combined products. ATC Classification system is provided by the World Health Organization to classify active substances in a hierarchy with five different levels [3]. *The system has fourteen main anatomical/pharmacological groups or 1st levels. Each ATC main group is divided into 2nd levels which could be either pharmacological or therapeutic groups. The 3rd and 4th levels are chemical, pharmacological or therapeutic subgroups and the 5th level is the chemical substance. The 2nd, 3rd 5th and 4th levels are often used to identify pharmacological subgroups when that is considered more appropriate than therapeutic or chemical subgroups.*

Drug information retrieval has been investigated in recent years. Adam S. Brown et al. have worked on information retrieval to design a database for drug repositioning [4]. Martin Kralinger et al. have worked on information retrieval for Chemistry, including drugs, using text mining methods [5]. Serkan Ayvaz et al. have worked on information retrieval about drug-drug interaction from publicly available sources [6].

Our paper aims at providing a knowledge-based automated tool to identify combination drugs from information provided onto a public database in the domain of antibiotherapy. Section 2 will describe methods. Results will be presented in section 3 and discussed in section 4.

2. Methods

Our method used to distinguish combined and plain pharmaceutical product uses a bank of successive filters. In our case, four consecutive filters have been designed and implemented.

All pharmaceutical specialties are listed in a file called *CIS bdpm.txt*. Each pharmaceutical specialty is identified by a code called CIS. This file allows to constitute the starting list of products from which we want to extract and isolate combined products. It contains 14 906 specialties.

2.1. Homeopathic medicines filter

According to World Health Organization, homeopathy is a system of medicine born in Europe in the last part of the eighteenth century. In some cases, the dilution is so high that it is almost impossible to find one molecule of the original raw material [7].

By definition, homeopathic medicines may contain several active ingredients. They are referenced, in the Summary of Product Characteristics as homeopathic medicines, or as medicines dedicated to homeopathy. As a consequence, searching for the terms

"homeopathy" or "homeopathic" in the Summary of Product Characteristics HTML pages allows to identify homeopathic medicines.

2.2. Antibiotics filter

Antibiotics are recognized from their "J01" ATC Class. ATC code is given in the Summary of Product Characteristics in HTML pages, which can be obtained by using a regular expression and basic text processing methods. Regular expressions allow to search for a string in a text document by describing their character composition. The ATC code is described as a seven characters string that starts by a capital letter, followed by two numbers, followed by two capital letters followed by two numbers. We observed that some ATC codes used in the HTML pages were misspelled using the O character instead of the 0 number. We integrated this possible misspelling in our regular expression, and used a post-processing step to correct the found ATC Code.

Antibiotics were defined to belong to the "J01" ATC Class. It means that their ATC Code starts with "J01" string.

2.3. Brand vs generic product filter

An interesting information is to be able to distinguish between the brand products and its generics. This information is provided in a text file called *CIS_GENER_bdpm.txt*. Generic groups are defined, identified by a code and a label. For each specialty, identified by a CIS code, the associated generic groups is given; it is also mentioned whether the drug product is an originator brand, a generic, a generic by dosage complementarity, or a substitutable generic. The aim of this filter is to select from a start list of specialties those that are a kind of generic product.

2.4. Single-substance filter

The entire composition of a specialty is listed in a file called *CIS_COMPO_bdpm.txt*. For each specialty, identified by its CIS code, the different pharmaceutical forms are listed separately. For each of them, the whole composition is detailed. Each substance is identified by an identifier, and tagged as *active substance* or *therapeutic fraction*. Indeed, the therapeutic fraction is defined as part of an active substance that carries pharmacological activity. Hence, the dose of the therapeutic fraction is different from the dose of the active substance, and both are provided in the file in two different lines, with a key value allowing to associate corresponding *active substance* and *therapeutic fraction*. If the *therapeutic fraction* is the same as the *active substance*, then only the information about the *active substance* is provided.

The aim of this filter is to remove from the initial list all specialties that have only one *active substance* in their composition. To perform that, therapeutic fractions should not be considered.

3. Results

Our method was evaluated to filter all specialties present in the *CIS_bdpm.txt* file.

Table 1 gives the number of specialties remaining after using each of the filters defined in the previous section. The initial list contained 14 906 specialties. From all there entities, we aim at finding those that are combined products. A chain of combined filters is used for each test; number and percentage of specialties remaining after the filters is provided in the second and third column of the table.

Table 1. Results of the different filters, used separately or combined to each other

Combination of used Filters	Number of remaining specialties	Percentage of removed specialties
Single substance filter	2 245	84.94%
Antibiotics filter	779	94.77 %
Homeopathic medicines filter	14 806	0.67%
Originator brand filter	1 261	91.54%
Single substance + Antibiotics filters	123	99.17 %
Single substance + Homeopathic medicines filters	2 161	85.50%
Single substance + originator brand product filter	176	98.82%
Antibiotics + Homeopathic medicines filters	779	94.77%
Single substance + Antibiotics + homeopathic medicine filters	123	99.17%
Single substance + Antibiotics + homeopathic medicine + originator brand product filters	11	99.93%

As we see, filters can be combined to retrieve the desired drug. From this drug associated characteristics can also be retrieved from the structured text files, like the price, or the reimbursement rate. Natural language processing methods can also be used to retrieve other data like the drug schema, or the contraindications.

4. Discussion

We proposed a method for retrieving combined antibiotic drug based on various filters. The method allows to considerably reduce the number of specialties, and can easily be extended to other types of drugs. Information provided in structured files of the French public drug database is limited, which may prevent from exploiting it in a clinical use purpose. For example, ATC codes, drug schema, contraindications, adverse reactions, excipients, and information about pregnancy and breast feeding are not provided in the structured files. Hence, they should be retrieved from the HTML pages using web connection and natural language processing methods. Web connection is time consuming and makes it very long a mass processing of a database. Natural language processing method are highly dependent of the quality of the data [8].

Moreover, the quality of the HTML pages is also limited. Instead of nesting HTML tags hierarchically, they follow each other in a listed way, making it more difficult to work with section header identifiers. In addition, information may be wrong. For example, we observed that the ATC code was written using the “O” character instead of the “0” number. Once it has been retrieved, it needs to be corrected to be used. For some specialties, the ATC code was absent from the web pages, and it was thus impossible to exploit it to generated and infer other data and information.

Finally, we observed that structured files were inconsistent. For instance, 7 497, i.e. 50.30% of all specialties, were absent of the structured file providing generic groups information. For those specialties, we were not able to infer any result related to the generic groups.

Identifying combined drug products could facilitate the design of decision support systems for alerting drug-drug interactions [9,10].

Acknowledgement

Funding was obtained from the Agence Nationale de Sécurité du Médicament et des Produits de santé (ANSM). APP 2016—RaMiPA Project (Raisonner pour Mieux Prescrire les Antibiotiques [Reasoning for a better antibiotic prescription]).

References

- [1] R. Kheshti, M. Aalipour, S. Namazi1, A comparison of five common drug–drug interaction software programs regarding accuracy and comprehensiveness, *Journal of Research in Pharmacy Practice* **5(4)** (2016), 257-263
- [2] <http://base-donnees-publique.medicaments.gouv.fr/>
- [3] https://www.whooc.no/atc_ddd_index/
- [4] A.S. Brown, C.J. Patel, A standard database for drug repositioning, *Scientific Data* volume **4** (2017), 170029
- [5] M. Krallinger, O. Rabal, A. Lourenço, J. Oyarzabal, A. Valencia, Information Retrieval and Text Mining Technologies for Chemistry, *Chemical Reviews* **117(12)**, 7673-7761
- [6] S. Ayvaz, J. Horn, O. Hassanzadeh, Q. Zhu, J. Stan, N.P. Tatonetti, S. Vilar, M. Brochhausen, M. Samwald, M. Rastegar-Mojarad, M. Dumontier, R.D. Boycek, Toward a complete dataset of drug–drug interaction information from publicly available sources, *Journal of Biomedical Informatics* **55** (2015),
- [7] <https://www.who.int/medicines/areas/traditional/Homeopathy.pdf>
- [8] D.S. Carrell, R.E. Schoen, D.A. Leffler, M. Morris, S. Rose, A. Baer, S.D. Crockett, R.A. Gourevitch, K.M. Dean, A. Mehrotra, Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings, *Journal of Medical Informatics Association* **24(5)**, (2017), 98 6-991, .
- [9] C.L. Tolley, S.P. Slight, A.K. Husband, N. Watson, D.W. Bates, Improving medication-related clinical decision support, *American Journal of health-system pharmacy* **75(4)** (2018), 239-246
- [10] H. Tilson, L.E. Hines, G. McEvoy, D.M. Weinstein, P.D. Hansten, K. Matuszewski, M. le Comte, S. Higby-Baker, J.T. Hanlon, L. Pezzullo, K. Vieson, A.L. Helwig, S.M. Huang, A. Perre, D.W. Bates, J. Poikonen, M.A. Wittie, A.J. Grizzle, M. Brown, D.C. Malone, Recommendations for selecting drug-drug interactions for clinical decision support, *American Journal of health-system pharmacy* **73(8)** (2016), 576-585

Better Safe than Sorry - Implementing Reliable Health Data Anonymization

Raffael BILD^{a,1}, Klaus A. KUHN^a and Fabian PRASSER^{b,c}

^aUniversity hospital rechts der Isar, Technical University of Munich, Germany

^bCharité – Universitätsmedizin Berlin, Berlin, Germany

^cBerlin Institute of Health (BIH), Berlin, Germany

Abstract. Modern biomedical research is increasingly data-driven. To create the required big datasets, health data needs to be shared or reused, which often leads to privacy challenges. Data anonymization is an important protection method where data is transformed such that privacy guarantees can be provided according to formal models. For applications in practice, anonymization methods need to be integrated into scalable and reliable tools. In this work, we tackle the problem of achieving reliability. Privacy models often involve mathematical definitions using real numbers which are typically approximated using floating-point numbers when implemented as software. We study the effect on the privacy guarantees provided and present a reliable computing framework based on fractional and interval arithmetic for improving the reliability of implementations. Extensive evaluations demonstrate that reliable data anonymization is practical and that it can be achieved with minor impacts on executions times and data utility.

Keywords. data protection, anonymization, reliable computing

1. Introduction

Modern data-driven biomedical research, e.g. in the field of precision medicine which tailors healthcare to the characteristics of individuals, increasingly leverages data science methods such as machine learning [1]. However, when creating the required big datasets, laws and regulations mandate stringent privacy protection. Hence, a wide range of safeguards has to be applied, ranging from organizational to technical measures.

Data anonymization is an important technical building block for implementing privacy protection. In this process, data is transformed in such a manner that formal guarantees, e.g. regarding the risk of singling out, linkage or inference, can be provided. Traditional models of privacy protection such as k -anonymity, ℓ -diversity and t -closeness specify syntactic constraints on output datasets, while more recent models like differential privacy formulate requirements for data processing methods [2].

2. Objective

All methods for implementing privacy models require performing changes to data which inevitably leads to a decrease of its utility. To balance a decrease in privacy risks with a

¹Corresponding Author: Raffael Bild, Institute of Medical Informatics, Statistics and Epidemiology, School of Medicine, Technical University of Munich, Ismaninger Straße 22, 81675 Munich, Germany; E-mail: raffael.bild@tum.de.

decrease of utility, models for quantifying both aspects have been developed. When implementing privacy models in practice, an important challenge lies in the need to reflect their mathematical definitions in software. Privacy models are often formulated over real numbers, which in software are approximated by floating-point numbers with limited precision (typically 64 bits). Computations can therefore yield results that differ significantly from the exact mathematical results [3]. This can make output data of anonymization tools vulnerable to privacy breaches. For example, it has been shown that straightforward implementations of a common method for achieving differential privacy can be exploited to extract the entire content of a (presumably protected) dataset [4]. However, studies of the effects of floating-point errors on the privacy guarantees provided by other methods for data anonymization are still lacking in the literature.

In this article, we aim to fill this gap, with a focus on investigating further methods which are truthful (i.e. non-perturbative) and hence particularly well-suited for the biomedical domain [5]. For this, we discuss the numerical stability of implementations of various privacy models, including k -anonymity, ℓ -diversity, t -closeness and further methods for achieving differential privacy [2]. Moreover, we present a reliable computing framework, which we have integrated into the open source data anonymization tool ARX [6] to mitigate vulnerabilities resulting from the use of floating-point operations.

3. Methods

3.1. Data Anonymization and Floating-Point Arithmetic

Figure 1 shows an example transformation of an input dataset using a combination of generalization (i.e. the replacement of values with more general, but semantically consistent values), suppression (i.e. the removal of values) and aggregation (i.e. the replacement of values with an aggregate, such as their mean). The example output dataset satisfies 2-anonymity, which means that each combination of attribute values appears at least twice (see [2] for further details). Whether or not k -anonymity is satisfied is easy to determine by simply counting the size of groups of indistinguishable records. Implementing other privacy models, such as ℓ -diversity, t -closeness or differential privacy, requires evaluating mathematical expressions over real numbers, though (cf. Section 3.2).

Age	Gender	Height	Age	Gender	Height
23	Male	176	[20-40[Male	179
35	Male	182	[20-40[Male	179
55	Male	176	[40-60[---	169
42	Female	162	[40-60[---	169

Generalization Suppression Aggregation

Figure 1. Example of input data and transformed output data.

In computers, real numbers are typically approximated using floating-point numbers. The number of floating-point values which can be represented with a fixed number of bits (typically 64) is finite. Hence, there exists an infinite number of unrepresentable real numbers. Most implementations of floating-point arithmetic adopt the IEEE standard 754 [7]. It specifies that all floating-point operations have to be performed as if it was possible to perform the corresponding operation with infinite precision, and then to round the result to a representable number. This inevitably introduces rounding errors which add up during sequences of calculations [3].

3.2. Numerical Stability of Common Privacy Models

Implementing some privacy models supported by ARX, e.g. k -anonymity [2], doesn't require decimal numbers at all. Implementing others requires significant amounts of decimal arithmetic, though. Examples are (1) t -closeness which basically requires that the distribution of sensitive attribute values in a set of indistinguishable data records is not too different from the corresponding distribution in the overall dataset or (2) (entropy) ℓ -diversity which requires the distribution (p_1, \dots, p_m) of sensitive attribute values in each group of indistinguishable records to satisfy $-\sum_{i=1}^m p_i \ln(p_i) \geq \ln(\ell)$ [2]. However, by studying possible effects of floating-point error propagation using forward analyses (see e.g. [3] for details), we were able to derive upper bounds for the resulting additive exceedances of the privacy parameters of these models. While a detailed presentation of these analyses is beyond the scope of this article, they showed that the resulting privacy violations are negligible in practice for all syntactic privacy models supported by ARX.

Differential privacy is not a property of a dataset, but a property of a data processing method. It basically guarantees that the probability of any possible output of a probabilistic algorithm does not change “by much” if data of one individual is added to or removed from the input dataset. The Laplacian mechanism and the exponential mechanism are important building blocks for implementing differentially private algorithms [8]. In [9], we have presented a process for implementing k -anonymity while fulfilling (ϵ, δ) -differential privacy. This approach uses random sampling to introduce non-determinism and the exponential mechanism to optimize the utility of output data. Consequently, unlike the majority of differentially private algorithms, it is truthful and therefore well-suited for processing health data [5].

We were able to calculate an upper bound on the rounding error induced by straightforward floating-point implementations of the exponential mechanism. For this, we applied conservative methodologies described e.g. in [3] followed by an extension of the original proof of the privacy guarantees provided [8] which takes rounding errors into account. While a detailed presentation of this analysis is, again, out of the scope of this article, it showed that the additive exceedance of the expected privacy loss ϵ is negligible, with values of about 10^{-10} or less in practical settings.

However, the implementation of differential privacy in ARX requires complex calculations to determine the sampling fraction β and the parameter k for k -anonymity to guarantee the requested degree of privacy protection. Investigating our floating-point implementation, we found that the deviations of ϵ were in the order of 10^{-16} using common values of ϵ (e.g. 0.01, 0.1, 0.5, $\ln(2)$, 0.75, 1, $\ln(3)$, 1.25, 1.5 and 2). The actual values calculated for the parameter δ , however, deviated drastically, as is shown in Table 1.

Table 1. Relative error of δ for $\epsilon = \ln(2)$ using a floating-point implementation.

Requested value of δ	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-8}	10^{-9}
Error of δ [%]	13.3	11.1	17.2	28.4	10.0	27.9

3.3. Design of ARX's Reliable Computing Framework

To solve this problem, we implemented a framework comprising different computing technologies that are reliable, i.e. offer strict guarantees for the accuracy of the calculated results: (1) Arithmetic using exact arbitrary-precision floating-point numbers. This can be used for calculations involving numbers with a finite amount of digits only. (2) Using representations as fractions with arbitrarily long integer enumerators and denominators. This approach can be used to perform exact calculations over rational numbers but it can

become very slow. (3) Interval arithmetic [3], which dynamically computes strict bounds on the errors of mathematical operations. The basic idea is not to operate on (approximations of) real numbers, but rather on intervals which enclose the respective exact real numbers. Functions operating on such intervals yield intervals which are guaranteed to include the exact result. For example, addition can be performed by calculating

$$[x_1, x_2] + [y_1, y_2] = [x_1 + y_1, x_2 + y_2].$$

ARX is implemented in Java and arbitrary prevision arithmetic and fraction arithmetic is well supported by common programming libraries. The number of Java libraries for performing interval arithmetic is, however, known to be limited [10]. Hence, we implemented a basic interval arithmetic framework from scratch while focusing on a minimal amount of easily understandable and verifiable code. We implemented various operators, including the basic arithmetic operators. For more complex functions such as *exp*, *pow*, *log* and *sqrt* we invoke the respective implementations for floating-point values provided by the Java standard library which have clearly defined accuracies. We also implemented various comparison operators such as \leq which allow for reliable comparisons by returning the result of comparing the upper and lower endpoint of their operands. These operators are guarded by checks which raise an error if the relation of their operands is not decidable, i.e. if the intervals are overlapping.

We used the methods and operators provided by this framework to implement the parameter calculation process for differential privacy reliably so that the actual degree of privacy protection provided can be at most more conservative than specified by the user.

4. Results

To evaluate the impact of the reliable parameter calculation on the strictness of the derived parameters we have compared it with a straight-forward floating-point implementation using common values of ϵ ranging from 0.01 to 2 and $\delta = 10^{-i}$ for $i = 1, \dots, 9$.

The differences between the values of β obtained using both implementations were very small with values of about 10^{-16} in all cases. All values obtained for k were identical except for ten configurations using irrational parameters. This is because these numbers have more significant figures than the other values considered, which resulted in higher rounding errors and hence larger intervals during calculations. In these cases, the values of k obtained reliably were (slightly) higher. Using $\epsilon = \ln(3)$, the values of k computed differed for $\delta = 10^{-5}$ and $\delta = 10^{-6}$ ($k = 63$ vs. $k = 66$ and $k = 78$ vs. $k = 82$, respectively). The results obtained when using $\epsilon = \ln(2)$ are listed in Table 2. As can be seen, the absolute differences were at most two. Consequently, for decreasing values of δ , which correspond to increasing degrees of privacy protection, the relative differences between the values of k obtained by both implementations tended to become smaller.

Table 2. Values of k derived from various values of δ and $\epsilon = \ln(2)$.

δ	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
Floating-Point	7	18	30	42	54	67	81	93	105
Reliable	8	20	32	44	56	68	81	95	107

In contrast to results obtained using the floating-point implementation (cf. Table 1), the actual values of ϵ and δ resulting from reliably calculated parameters k and β were at most more conservative than the privacy parameters specified by the user. In particular, increasing k was necessary to mitigate the violations of δ reported in Table 1. At the same time, the impacts on the intensity of data transformations applied and hence the potential

reductions of data utility are negligible when using recommended parameterizations [9].

We also evaluated the execution times of both implementations on a PC with a quad-core 3.1 GHz CPU, Ubuntu Linux and an Oracle JVM. The results are shown in Figure 2. When decreasing both ϵ and δ (which corresponds to stronger degrees of protection), the relative execution times tended to increase. Using typical values of $\epsilon \approx 1$ and $\delta \approx 10^{-6}$, the execution time of the reliable implementation was about four times higher than the time used by the floating-point implementation. In all experiments with $\epsilon \geq 0.1$, the calculation of parameters terminated in less than one second using both implementations. This contains the range of parameters which is practical for the approach (cf. [9]).

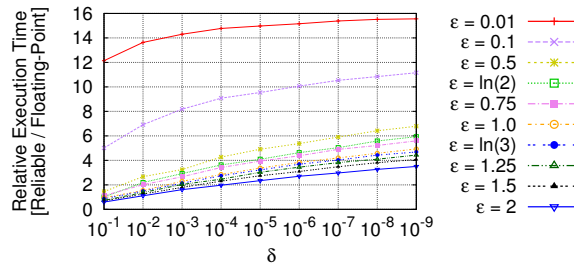


Figure 2. Execution times for deriving β and k from various values of δ and ϵ reliably relative to the floating-point implementation.

5. Conclusion

In this article, we studied how privacy violations resulting from floating-point implementations of anonymization algorithms can be mitigated. We discussed reliability issues resulting from arithmetic operations for a variety of privacy models including k -anonymity, ℓ -diversity and t -closeness as well as an implementation of differential privacy specifically suited for applications to health data [9]. Moreover, we presented a framework comprising reliable computing techniques, including interval and fractional arithmetic. All results have been integrated into the open source tool ARX. Finally, we examined the impacts of the reliable implementation on output data utility as well as execution times and found both to be negligible in practice when realistic parameters are being used.

References

- [1] V. Gligorić et al., Integrative methods for analyzing big data in precision medicine, *Proteomics* **16** (2016), 741–758.
- [2] B. C. Fung et al., *Introduction to privacy-preserving data publishing: Concepts and techniques*, CRC Press, 2010.
- [3] N. J. Higham, *Accuracy and stability of numerical algorithms*, vol. 80, Siam, 2002.
- [4] I. Mironov, On significance of the least significant bits for differential privacy, *Proceedings of the 2012 ACM conference on computer and communications security*, ACM, 2012, 650–661.
- [5] F. K. Dankar and K. El Emam, Practicing differential privacy in health care: A review., *Transactions on data privacy* **6** (2013), 35–67.
- [6] F. Prasser and F. Kohlmayer, Putting statistical disclosure control into practice: The ARX data anonymization tool, *Medical data privacy handbook*, Springer, 2015, 111–148.
- [7] IEEE Standards Committee et al., 754-2008 IEEE standard for floating-point arithmetic, *IEEE computer society std* **2008**.
- [8] C. Dwork, A. Roth et al., The algorithmic foundations of differential privacy, *Foundations and trends® in theoretical computer science* **9** (2014), 211–407.
- [9] R. Bild, K. A. Kuhn and F. Prasser, Safepub: A truthful data anonymization algorithm with strong privacy guarantees, *Proceedings on privacy enhancing technologies* **2018** (2018), 67–87.
- [10] J. W. von Gudenberg, OOP and interval arithmetic—Language support and libraries, *Numerical software with result verification*, Springer, 2004, 1–14.

Blood Lactate Concentration Prediction in Critical Care

Behrooz MAMANDIPOOR^{1 a}, Mahshid MAJD^a, Monica MOZ^{a,b} and Venet OSMANI^a

^a *Fondazione Bruno Kessler Research Institute, Trento, Italy*

^b *Humanitas Research Hospital, Milan, Italy*

Abstract. Blood lactate concentration is a reliable risk indicator of deterioration in critical care requiring frequent blood sampling. However, lactate measurement is an invasive procedure that can increase risk of infections. Yet there is no clinical consensus on the frequency of measurements. In response we investigate whether machine learning algorithms can be used to predict blood lactate concentration from ICU health records. We evaluate the performance of different prediction algorithms using a multi-centre critical care dataset containing 13,464 patients. Furthermore, we analyse impact of missing value handling methods in prediction performance for each algorithm. Our experimental analysis show promising results, establishing a baseline for further investigation into this problem.

Keywords. machine learning, deep learning, clinical decision support, critical care

1. Introduction

Blood lactate concentration is frequently measured in critical care as it is used as an indicator of risk of patients' deterioration. Presence of increased lactate levels is correlated with increased risk of morbidity and mortality. In healthy individuals there is a continuous cycle of lactate production and clearance (metabolised primarily in the liver), while in critically ill patients lactate metabolism is impaired, resulting in elevated lactate concentration. Several clinical conditions have been associated with impaired clearance of lactate, such as liver dysfunction and sepsis. Therefore frequent lactate measurements are necessary to track and assess patients' state. However, lactate concentration cannot be measured without drawing arterial or venous blood, which is an invasive procedure that can increase risk of infections. As a consequence, blood gas analysis may not be ordered as frequently leading to sub optimal rates of lactate measurements [1].

In this respect, there is ample potential for machine learning methods to play a significant role in lactate guided clinical decision making. Such role is especially important when considering that lactate-guided therapy significantly reduced hospital mortality and length of stay as evidenced by several multicentre randomised controlled trials [2,3,4]. However, this potential has not been explored thus far. Addressing the evident need for prediction of blood lactate concentration in ICU, we are the first to compare performance

¹ Corresponding Author: E-mail: bmamandipoor@fbk.eu

of several machine learning methods in lactate concentration prediction. Furthermore, we evaluate several strategies to handle missing values using a dataset of 13,464 patients, containing 12,196,798 clinical records from the eICU critical care database [5]. Our results show that blood lactate concentration prediction could become a useful tool during clinical decision making process in critical care, that may reduce unnecessary blood sampling, while meeting recommendations of Surviving Sepsis Campaign (SSC) on serial lactate testing [6].

2. Methods

We formally define the problem of lactate concentration prediction as follows: For each patient with the set of clinical parameters $S = \{(l, X)_t; t \in T\}$, where l represents blood lactate concentration, X represents the set of all other clinical measures, and t is the time index, we want to predict $l_{t+\beta}$ based on $\hat{S}_t = \{l, \hat{X}\}_{t'} \in [t - \alpha: t]$, where \hat{X} is the selected set of clinically relevant measurements out of X . Therefore, we formulate a regression problem using Eq.(1).

$$f: \hat{S}_t \rightarrow l_{t+\beta}, \min_{\theta} \text{Loss}(f(\hat{S}_t, \theta), l_{t+\beta}) \quad (1)$$

In other words, the objective of lactate prediction is to predict the blood lactate concentration of a patient in the next β hours using a selected set of their measurements taken in the past α hours.

For our work we use eICU Collaborative Research Database [5], a multi-center intensive care unit database with high granularity data. The eICU database comprises 200,859 patient unit encounters for 139,367 unique patients admitted between 2014 and 2015 to hospitals located throughout the US. The final patient cohort contained 13,464 patients (14,477 ICU stays) with 12,196,798 clinical records, where we grouped these records into 2 hour windows. Patient's mean age was 61.8 ± 15.7 years (45% female). For our case study, the relevant clinical features for the selected cohort are chosen based on advice from the clinician and outlined in Table 1.

Critical care data suffers from several limitations with missing data as one of the main challenges. In the eICU dataset, the percentage of missingness per feature is between 0 to 96 which needs to be addressed properly. Addressing missing values is typically dependent on the cause of missingness. Statistically speaking, there are three types of missing values: missing completely at random (MCAR) which happens on an unrelated cause, missing at random (MAR) which implies a relation between missing value and other present values, and missing not at random (MNAR) which implies a relationship between the value of the variable and its missingness [7]. On the other hand, there are several imputation methods to handle missingness. Mean and forward imputations are basic single-value imputation methods, which only consider the information in the past values of the variable and work best under MCAR assumption. Multiple imputation [8], Matrix Factorisation [9], PCA [10], SoftImpute [11] and Random Forest [12] are the most known traditional machine learning methods that find the substitute for missing values based on the relation between observed and missing features. These methods perform best when handling MAR cases. Finally, MNAR cases are the hardest to manage. A helpful solution to capture the information behind missingness of variables is to use missing

Table 1. List of selected variables from eICU tables (in bold) based on clinical relevance

Patient , gender, age, ethnicity, admissionweight, apacheadmissiondx
Lab , Respiratory Rate, O2 Saturation, FiO2, glucose, potassium, sodium, Hgb, chloride, creatinine, BUN, bicarbonate, LPM O2, calcium, Hct, platelets x 1000, anion gap, WBC x 1000, lactate, RBC, RDW, paO2, pH, paCO2, magnesium, HCO3, Total CO2, Base Excess, phosphate, Pressure Support, albumin, -lymphs, -polys, -eos, -basos, -bands, -monos, bilirubin, AST (SGOT), ALT (SGPT), total protein, alkaline phos.
NurseCharting , Heart Rate, Respiratory Rate, Temperature (C), Invasive BP Mean, Invasive BP Systolic, Invasive BP Diastolic, Non-Invasive BP Mean, Non-Invasive BP Systolic, Non-Invasive BP Diastolic, O2 Saturation, glucose, CVP, LPM O2, Total CO2
RespiratoryCharting , Heart Rate, Respiratory Rate, FiO2, Total CO2, Tidal Volume, Inspiratory Pressure, LPM O2, Vent Rate, Plateau Pressure, Mean Airway Pressure, Pressure Support, Inspiratory Pressure
VitalPeriodic , Heart Rate, Respiratory Rate, Temperature (C), Invasive BP Mean, Invasive BP Systolic, Invasive BP Diastolic, CVP
VitalAperiodic , Non-Invasive BP Mean, Non-Invasive BP Systolic, Non-Invasive BP Diastolic
IntakeOutput , Urine

indicators. These indicators can either be used directly aside data or inside the prediction model structure. Note that indicators can also be useful in MAR and MCAR, since they differentiate the imputed values from the observed ones and the prediction method can use this information to ignore bad imputations.

There are a vast number of successful regression methods used for healthcare data which could be grouped into three categories: the statistical regression methods, the traditional machine learning methods and the rising deep learning methods. Here we investigate one method from each category: Lasso regression(LR) as linear regression, to benefit from its feature selection capabilities, Random Forest(RF) as nonlinear model with an ensemble of decision trees, and Long Short-term Memory(LSTM) as temporal deep learning model to learn complex temporal relation in data.

3. Results

Data preparation. Data is preprocessed in six steps. (1) Patient cohort is selected based on three inclusion criteria; adult patients (age > 18), with at least two measured lactate levels, and with at least 18 hours length of stay in ICU. (2) The relevant variables are selected based on advice from the clinician. For our case study, the relevant clinical features for the selected cohort are outlined in Table 1. A detailed description of each variable is provided in the original eICU paper [5]. (3) Some features exist in more than one table under different names. These features are aligned to a unique feature in time. (4) The selected data is aligned in time since each feature is measured in an arbitrary time and frequency. We resampled time-series data into regularly aligned periods where each feature is sampled every 2 hours. In case a feature is measured more than once during each two hour interval, the last record is used. (5) Noise and outliers are addressed as follows: for each feature, the valid interval is defined based on clinical knowledge and the values out of the valid scope are considered as missing values. (6) The data is split into training and testing parts, which is done using five-fold cross validation where at each fold 80% of the data is considered as training and the rest is test data.

Settings. All the imputation methods are applied using python based on

fancyimpute, predictive imputer, and sklearn open-source libraries. Also for all the machine learning methods, the default parameters proposed by their authors are used. The LSTM network has 2 layers of 1024 units with *Glorot* normalisation and *tanh* activation, each followed by a drop out layer of 0.6. Adam optimizer is used with learning rate starting from 0.0001 and the model is trained for 20 epochs with batch size 100. To ease LSTM convergence, data is normalised to have zero mean and a standard deviation of one. The LSTM model is implemented using Keras and Tensorflow as backend and one GTX 2080 Ti as GPU.

All combinations of introduced prediction models and imputation methods are examined on eICU data and their results are reported in Table 2. To measure the quality of imputation methods regarding lactate prediction while preserving the structure of data, we report Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) as indicators of the predictive performance of the regression models. The mean and standard deviation of each measure on five fold cross-validation of data is reported. The best results are shown in bold. As the results suggest, both regression model and imputation method affect the prediction results. LSTM performed significantly better compared to LR an RF using the same imputation method. Therefore, it can be concluded that the data contains complex relations not only between measurements but also across their values in time that were captured by the LSTM model.

Table 2. The results of all combinations of prediction models and imputation methods. The mean and standard deviation of various metrics on five fold cross validation of data is provided. MAE, RMSE, and R^2 stand for Mean Absolute Error, Root Mean Squared Error, and R-squared respectively. LR, RF, and LSTM stand for Linear Regression, Random Forest, and Long Short Term Memory.

Measure	Regression Imputation		LR	RF	LSTM
MAE	Mean		0.859 ± 0.006	0.856 ± 0.007	0.745 ± 0.045
	Feed Forward		0.735 ± 0.008	0.733 ± 0.009	0.692 ± 0.008
	Indicator		0.725 ± 0.009	0.720 ± 0.010	0.665 ± 0.009
	PCA		0.864 ± 0.006	0.861 ± 0.008	0.712 ± 0.008
	MF		0.905 ± 0.009	0.901 ± 0.010	0.715 ± 0.012
	SoftImpute		0.862 ± 0.010	0.858 ± 0.011	0.705 ± 0.006
	MissForest		0.853 ± 0.008	0.849 ± 0.005	0.714 ± 0.010
	MICE		0.872 ± 0.007	0.869 ± 0.008	0.715 ± 0.009
	AE		0.846 ± 0.007	0.845 ± 0.008	0.730 ± 0.051
RMSE	Mean		1.263 ± 0.013	1.257 ± 0.015	1.120 ± 0.016
	Feed Forward		1.120 ± 0.014	1.115 ± 0.015	1.075 ± 0.016
	Indicator		1.090 ± 0.013	1.085 ± 0.016	1.016 ± 0.025
	PCA		1.268 ± 0.014	1.262 ± 0.016	1.104 ± 0.014
	MF		1.304 ± 0.015	1.298 ± 0.016	1.100 ± 0.018
	SoftImpute		1.268 ± 0.016	1.264 ± 0.017	1.095 ± 0.017
	MissForest		1.254 ± 0.016	1.248 ± 0.011	1.100 ± 0.011
	MICE		1.270 ± 0.014	1.266 ± 0.015	1.105 ± 0.016
	AE		1.245 ± 0.013	1.241 ± 0.015	1.103 ± 0.022
R^2	Mean		0.475 ± 0.010	0.480 ± 0.013	0.585 ± 0.022
	Feed Forward		0.587 ± 0.008	0.591 ± 0.011	0.620 ± 0.013
	Indicator		0.605 ± 0.012	0.610 ± 0.013	0.660 ± 0.017
	PCA		0.471 ± 0.011	0.476 ± 0.014	0.598 ± 0.011
	MF		0.440 ± 0.013	0.445 ± 0.015	0.601 ± 0.014
	SoftImpute		0.471 ± 0.010	0.475 ± 0.013	0.605 ± 0.006
	MissForest		0.482 ± 0.007	0.487 ± 0.006	0.601 ± 0.012
	MICE		0.469 ± 0.010	0.472 ± 0.012	0.598 ± 0.011
	AE		0.489 ± 0.010	0.493 ± 0.012	0.599 ± 0.024

4. Conclusion

Considering the importance of monitoring blood lactate levels, the ability to predict blood lactate concentration may provide clinicians sufficient time to devise interventions for rapidly deteriorating patients. At the same time it may decrease the frequency (and the associated cost) of invasive lactate measurements in stable patients. Therefore, we defined the problem of lactate prediction from ICU data in details and offered solutions with well-known prediction and imputation methods.

Our results show that LSTM-based method can predict lactate level with a Mean Absolute Error of 0.665 across 13,464 patients from different hospitals and ICU units. Furthermore, we show that indicator imputation method achieves highest performance in our dataset, suggesting that a missing value indicator is informational and increases predictive power over other, mean-based imputation methods. Future work will investigate prediction of lactate levels in patient sub-populations and its impact on critical care at patient as well as organisational level.

This work was partially supported by the European Union's Horizon 2020 research and innovation program under grant agreement #769765.

References

- [1] C. Rhee, M. V. Murphy, L. Li, R. Platt, M. Klompas, Lactate testing in suspected sepsis: trends and predictors of failure to measure levels, *Critical care medicine* 43 (8) (2015) 1669 (2015).
- [2] T. C. Jansen, J. van Bommel, F. J. Schoonderbeek, S. J. Sleeswijk Visser, J. M. van der Klooster, A. P. Lima, S. P. Willemsen, J. Bakker, Early lactate-guided therapy in intensive care unit patients: a multicenter, open-label, randomized controlled trial, *American journal of respiratory and critical care medicine* 182 (6) (2010) 752–761 (2010).
- [3] A. E. Jones, N. I. Shapiro, S. Trzeciak, R. C. Arnold, H. A. Claremont, J. A. Kline, E. M. S. R. N. E. Investigators, et al., Lactate clearance vs central venous oxygen saturation as goals of early sepsis therapy: a randomized clinical trial, *Jama* 303 (8) (2010) 739–746 (2010).
- [4] H. Tian, S. Han, C. Lv, T. Wang, Z. Li, D. Hao, Q. Shang, X. Wang, The effect of early goal lactate clearance rate on the outcome of septic shock patients with severe pneumonia, *Zhongguo wei zhong bing ji jiu yi xue*= Chinese critical care medicine= *Zhongguo weizhongbing jijiuyixue* 24 (1) (2012) 42–45 (2012).
- [5] T. J. Pollard, A. E. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, O. Badawi, The eicu collaborative research database, a freely available multi-center database for critical care research, *Scientific data* 5 (2018).
- [6] M. M. Levy, L. E. Evans, A. Rhodes, The surviving sepsis campaign bundle: 2018 update, *Intensive care medicine* 44 (6) (2018) 925–928 (2018).
- [7] D. B. Rubin, Inference and missing data, *Biometrika* 63 (3) (1976) 581–592 (1976).
- [8] M. J. Azur, E. A. Stuart, C. Frangakis, P. J. Leaf, Multiple imputation by chained equations: what is it and how does it work?, *International journal of methods in psychiatric research* 20 (1) (2011) 40–49 (2011).
- [9] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, *Computer* (8) (2009) 30–37 (2009).
- [10] J. Josse, F. Husson, Handling missing values in exploratory multivariate data analysis methods, *Journal de la Société Française de Statistique* 153 (2) (2012) 79–99 (2012).
- [11] R. Mazumder, T. Hastie, R. Tibshirani, Spectral regularization algorithms for learning large incomplete matrices, *Journal of machine learning research* 11 (Aug) (2010) 2287–2322 (2010).
- [12] D. J. Stekhoven, P. Bühlmann, Missforest—non-parametric missing value imputation for mixed-type data, *Bioinformatics* 28 (1) (2011) 112–118 (2011).

Building an I2B2-Based Population Repository for Clinical Research

Lydia GONZÁLEZ^a, David PÉREZ-REY^a, Enrique ALONSO^a, Gema HERNÁNDEZ^a,
Pablo SERRANO^b, Miguel PEDRERA^b, Agustín GÓMEZ^b, Kristof DE SCHEPPER^c,
Titus CREPAIN^c, Brecht CLAERHOUT^c

^a*Biomedical Informatics Group, Artificial Intelligence Department, E.T.S.I. Informáticos, Universidad Politécnica de Madrid, Spain.*

^b*Hospital Universitario 12 de Octubre, Madrid, Spain.*

^c*Custodix N.V., Sint-Martens-Latem, Belgium.*

Abstract. The present work provides a real-world case of the connection process of a hospital, 12 de Octubre University Hospital in Spain, to the TriNetX research network, transforming a compilation of disparate sources into a single harmonized repository which is automatically refreshed every day. It describes the different integration phases: terminology core datasets, specialized sources and eventually automatic refreshment. It also explains the work performed on semantic normalization of the involved clinical terminologies; as well as the resulting benefits the InSite platform services have enabled in the form of research opportunities for the hospital.

Keywords. I2B2, ICD, LOINC, ATC, SNOMED-CT, Clinical research, Secondary use of EHR, Semantic interoperability, Clinical data integration.

1. Introduction

The adoption of Electronic Health Record (EHR) systems for storing clinical data has been trending upwards in the last decades, especially in larger hospitals where record tracking has become increasingly difficult as they grow. However, the formats used by each department or area for recording and storing these data are rarely consistent. The process of extracting and combining information from several sources within a hospital in order to, for instance, allow observational studies on broad cohorts of patients, can be anything but trivial.

This scenario is where resources such as the InSite platform [1] come into play. InSite is a secure service platform which enables healthcare organizations across Europe to harmonize their clinical data sources in order to increase data availability and its efficient reuse for research. InSite, as the European branch of TriNetX [2], has helped establish a global Clinical Research Network, with over 130 healthcare providers across 23 countries and almost 4.200 clinical trial opportunities already delivered.

When a new hospital joins the network, InSite integrates its clinical sources to a local node, an I2B2 [3] repository where all this data is homogenized, which is linked to the central InSite platform. Other common data models such as OMOP or PCORnet could have been chosen as well, but I2B2 was selected for being a flexible and widely used standard [4, 5, 6]. Federated analytics technology ensures that patient data never

leaves the hospital and remains under local control, while providing aggregated results to research questions.

The present paper provides a real example of the complete connection process of 12 de Octubre University Hospital [7] in Spain to the InSite network, transforming a compilation of disparate sources into a single harmonized repository which is automatically refreshed on a daily basis.

2. Methods

During the data loading process for new hospitals, integration is not performed simultaneously for all sources. It is divided in iterations or phases: a historical data load, integration of specialized sources and lastly automation of incremental updates.

In the case of 12 de Octubre, the complete process required around 2 years of work, starting April 2017. Details on the specific steps taken and data sources involved in each of the phases are presented below.

2.1. Phase I – Minimum Basic Data Set (MBDS)

The initial phase consisted on a historical data load of the essential domains from the EHR system of the hospital (HP-HCIS, [8]). This includes Demographics (age, gender and vital status related data), Providers (e.g. physicians, nurses, etc.), Visits (any clinical transaction generating a set of data), Diagnoses, Procedures, Laboratory tests and Medication. This data was extracted from the source database and transformed, directly through series of queries, to fit the I2B2 structure of its target table.

When local coding systems do not match the reference terminologies used in InSite, specific mappings have to be developed and loaded into the hospital's I2B2 node in this step. The mapping effort to standard clinical terminologies has been performed by a multidisciplinary team from both 12 de Octubre and InSite.

For diagnoses and procedures, the hospital had generated mappings from their local coding to ICD9CM [9] or, for recent data, to ICD10CM and ICD10PCS [10]. In the interest of a joint exploitation of both historical and recent information in a single terminology, codes stored in ICD9CM were also mapped to ICD10CM or PCS respectively by InSite, using internally developed mappings to increase reliability and maintain legacy data [11, 12].

Translation of most local laboratory tests into LOINC [13] was provided by the hospital. For those local codes not present in the hospital's internal mapping, InSite developed a manual mapping to LOINC.

Finally, for medication data, AEMPS (*Spanish Agency of Medicines and Medical Devices*) codes were translated to ATC [14] using the official mapping [15], which was further refined by 12 de Octubre.

2.2. Phase II – Specialized Sources & Expansion of Previous Data

The second phase involved two main extensions of the harmonized data. Firstly, the integration of four new sources: cancer registry, anatomic pathology, clinical findings and oncological medication. Secondly, an update of the preexisting sources and the inclusion their additional visit types: inpatients, outpatients, ambulatory surgery or

emergency room. Two of the new sources were processed through exports of their original systems instead of direct database connections. These export files were then integrated with the rest of the collected information.

Regarding the terminologies assigned to the new sources, cancer registry and anatomic pathology data had been coded using a local terminology. Before entering the I2B2 repository, all topology and most morphology information was mapped by InSite to ICD-O-3 [16] while the remaining morphology information, together with specific cancer registry concepts such as TNM staging data, were mapped to SNOMED-CT [17].

Clinical findings data, similarly to what had been done for laboratory tests in phase I, was mapped automatically to SNOMED-CT through the hospital’s internal mapping tables before extraction. Oncological Medication was locally already in ATC.

2.3. Phase III – Data Load Automation

The first step taken during this phase of the integration process was to perform a load from all data sources in order to bring historical data completely up to date, bridging the time gap since the latest manual refreshment. This was followed by the introduction of automatic incremental updates, which begun by the implementation of HL7 (*Health Level 7*, [18]) channels for ingesting fresh data from the different sources.

The new loading process entails a two-step ETL. Whenever fresh data is recorded, individual HL7 messages are sent through a channel implemented by the hospital, connected to the source systems. Subsequently, these messages are processed in a channel implemented by InSite, which models the information in the format required by the I2B2 repository.

After each channel has been implemented, InSite continues to provide technical support and validation tasks as requested by the hospital to ensure its stability, such as filtering of duplicate messages, detection of empty fields or local unit errors.

There are currently eight channels ingesting messages periodically from the following data domains: diagnoses, procedures, laboratory tests, medication, oncology and clinical findings. These incremental updates occur daily. Fresh data extraction requires an average of 10 minutes, after which the steps of processing, quality checking and load take place. The I2B2 repository and InSite applications for the hospital’s internal use remain accessible at all times during this process.

3. Results

Tables 1 and 2 summarize the outcome of the integration process for phases I and II.

Table 1. Details on the historical data load of the Minimum Basic Data Set (MDBS) – Phase I.

Data domain	Visit type	Period	# Records	Target I2B2 Table
Demographics	Inpatients	–	2.672.821	PATIENT_DIMENSION
Providers	–	–	43.053	PROVIDER_DIMENSION
Visits	Inpatients	01/1999 - 02/2017	22.462.945	VISIT_DIMENSION
Diagnoses	Inpatients	01/2010 - 02/2017	1.990.697	OBSERVATION_FACT
Procedures	Inpatients	01/2010 - 02/2017	1.036.653	OBSERVATION_FACT
Laboratory tests	All	01/2016 - 02/2017	148.761.307	OBSERVATION_FACT
Medication	Inpatients	01/2011 - 02/2017	1.996.671	OBSERVATION_FACT

Table 2. Details on the load of specialized sources and additional visit types – Phase II. Visit types include Inpatients (IN), Outpatients (OUT), Ambulatory Surgery (SUR) and Emergency Room.

Data domain	Visit type	Period	# Records	Target I2B2 Table
Demographics	All	–	2.698.508	PATIENT_DIMENSION
Providers	–	–	43.700	PROVIDER_DIMENSION
Visits	–	01/1999 - 07/2017	23.579.551	VISIT_DIMENSION
Diagnoses	IN,OUT,SUR	01/2010 - 07/2017	3.142.581	OBSERVATION_FACT
Procedures	IN,SUR	01/2010 - 07/2017	1.730.628	OBSERVATION_FACT
Laboratory tests	All	01/2015 - 07/2017	226.872.641	OBSERVATION_FACT
Medication	IN,OUT,SUR	01/2008 - 07/2017	2.371.452	OBSERVATION_FACT
Cancer registry	All	01/1980 - 07/2017	66.966	OBSERVATION_FACT
Anat. pathology	All	01/2014 - 07/2017	783.894	OBSERVATION_FACT
Clinical find.	All	01/2010 - 07/2017	11.492.325	OBSERVATION_FACT
Oncology med.	All	01/2002 - 07/2018	252.267	OBSERVATION_FACT

After the implementation of incremental updates during phase III, the volume of harmonized data has continued to grow as fresh information is automatically included in the repository. Figure 1 illustrates the evolution of the volume per data domain over the integration process.

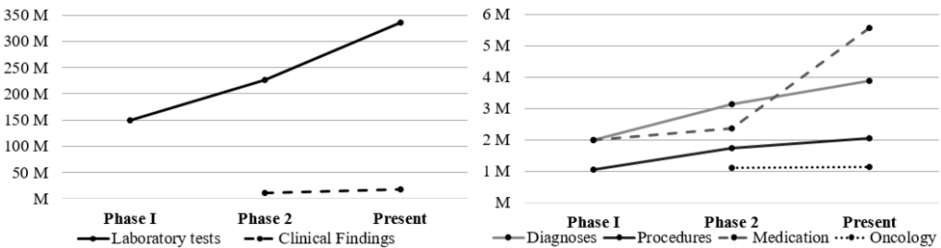


Figure 1. Snapshot of the data volume per domain through the different phases of the integration process.

Quality assurance processes were undertaken in parallel to the described ETL processes and a number of expected fluctuations have been identified in collaboration with the hospital’s IT team. As a future line of work, time series analyses of volume variations in the different data domains are being designed, in order to extract patterns of data peaks and drops, and to forecast its evolution for detecting unexpected variations.

As an example of this evolution of data volume over time, figure 2 shows the number of new laboratory records loaded per month in the period of January 2015 to August 2019. Indicator 1 marks the moment in which the hospital’s coding team implemented a filter to remove redundant laboratory records, thus reducing the mean monthly volume of new data. The second indicator marks the period in which laboratory data load migrated to the appropriate HL7 channel and automatic updating of this source began.



Figure 2. Monthly evolution of laboratory data. Grey: manually loaded. Black, dashed: with automatic updates.

4. Conclusions

The InSite team, in collaboration with the IT department of 12 de Octubre University hospital, have managed to integrate a number of heterogeneous, dissociated clinical data sources into a single, harmonized repository where information is refreshed daily, without manual intervention, on a record-to-record basis. Not only does this help the hospital's physicians access patient information from several data domains easily and efficiently, but thanks to incremental updates it also allows the hospital to perform clinical trial recruitment protocols with the least possible data availability delay. Keeping harmonized data up to date implies physicians can look for suitable patients including those currently under treatment.

Apart from the numerous clinical trial opportunities offered through the InSite platform, joining the network has enabled the participation of 12 de Octubre in a number of European research projects in collaboration with prominent healthcare and research institutions, including EHR2EDC[19], ECRIN[20], RIS[21], CDISC[22] or ICHOM[23].

Regarding future courses of action, complementary clinical information sources are continuously being identified and integrated within the hospitals of the InSite network. In the case of 12 de Octubre, new data sources currently considered are microbiology, genetics, HIV registry, medication posology and specialized parameters from breast and lung cancer. Further iterations of integration will provide essential data to drive the current clinical research more efficiently.

References

- [1] InSite, A TriNetX Company (2019). Retrieved from www.insiteplatform.com/
- [2] TriNetX (2019). Retrieved from www.trinetx.com/
- [3] I2B2, Informatics for Integrating Biology & the Bedside (2019). Retrieved from www.i2b2.org/
- [4] Weeks, J., & Pardee, R. (2019). Learning to share health care data: A brief timeline of influential common data models and distributed health data networks in US health care research. *eGEMS*, 7(1).
- [5] Waitman, L. R. et al. (2011). Expressing observations from electronic medical record flowsheets in an i2b2 based clinical data repository to support research and quality improvement. In *AMIA Annual Symposium Proceedings* (Vol. 2011, p. 1454). American Medical Informatics Association.
- [6] Hospital Universitario 12 de Octubre (2019). Retrieved from www.comunidad.madrid/hospital/12octubre/
- [7] Montero, M. A., & Prado, S. (2008). Electronic health record as a knowledge management tool in the scope of health. In *Workshop on Knowledge Management for Health Care Procedures* (pp. 152-166).
- [8] ICD, 9th Revision, Clinical Modification (2015). Retrieved from www.cdc.gov/nchs/icd/icd9cm.htm
- [9] ICD, 10th Revision (2016). Retrieved from www.cdc.gov/nchs/icd/icd10.htm
- [10] Hernández G. et al. (2019) ICD-10-CM extension with ICD-9 diagnosis codes to support integrated access to clinical legacy data. *International Journal of Medical Informatics*, 129, 189-197.
- [11] Hernández G. et al. (2019) ICD-10-PCS extension with ICD-9 procedure codes to support integrated access to clinical legacy data. *International Journal of Medical Informatics*, 122, 70-79.
- [12] Logical Observation Identifiers Names and Codes (2019). Retrieved from <https://loinc.org/>
- [13] ATC Classification System (2018). Retrieved from www.whocc.no/atc/structure_and_principles/
- [14] AEMPS Nomenclator (2017). Retrieved from cima.aemps.es/cima/publico/nomenclator.html
- [15] ICD-O, 3rd Ed. (2019). Retrieved from www.who.int/classifications/icd/adaptations/oncology/en/
- [16] SNOMED-CT (2019). Retrieved from www.snomed.org/snomed-ct/five-step-briefing
- [17] Health Level 7 International (2019). Retrieved from www.hl7.org
- [18] EHR2EDC (2018). Retrieved from www.eithealth.eu/ehr2edc
- [19] European Clinical Research Infrastructure Network (2019). Retrieved from www.ecrin.org
- [20] Red Española de Investigación en SIDA (2019). Retrieved from www.redris.es/inicio
- [21] Clinical Data Interchange Standards Consortium (2019). Retrieved from www.cdisc.org
- [22] International Consortium for Health Outcomes Measurement (2019). Retrieved from www.ichom.org

Character-Level Neural Language Modelling in the Clinical Domain

Markus KREUZTHALER^{a,b,1}, Michel OLEYNIK^a and Stefan SCHULZ^a

^a*Institute for Medical Informatics, Statistics and Documentation,
Medical University of Graz*

^b*CBmed GmbH — Center for Biomarker Research in Medicine*

Abstract. Word embeddings have become the predominant representation scheme on a token-level for various clinical natural language processing (NLP) tasks. More recently, character-level neural language models, exploiting recurrent neural networks, have again received attention, because they achieved similar performance against various NLP benchmarks. We investigated to what extent character-based language models can be applied to the clinical domain and whether they are able to capture reasonable lexical semantics using this maximally fine-grained representation scheme. We trained a long short-term memory network on an excerpt from a table of de-identified 50-character long problem list entries in German, each of which assigned to an ICD-10 code. We modelled the task as a time series of one-hot encoded single character inputs. After the training phase we accessed the top 10 most similar character-induced word embeddings related to a clinical concept via a nearest neighbour search and evaluated the expected interconnected semantics. Results showed that traceable semantics were captured on a syntactic level above single characters, addressing the idiosyncratic nature of clinical language. The results support recent work on general language modelling that raised the question whether token-based representation schemes are still necessary for specific NLP tasks.

Keywords. Neural Networks, Electronic Health Records, Natural Language Processing

1. Introduction and Motivation

So-called word embeddings have become popular for various natural language processing (NLP) tasks. They capture the semantics of a word to a certain degree and have replaced term-weighted vector-based representation schemes from the pre-deep learning era. For building such embeddings, a critical amount of textual data is needed to obtain stable vector representations with different surrounding contexts. The excellent performance of such representations, in combination with different neural network architectures were demonstrated in various clinical NLP tasks.

In most cases, character-based input representations were handled as a merely supportive information layer for word-level embeddings in different NLP problem domains.

¹Corresponding Author: Institute for Medical Informatics, Statistics and Documentation, Graz General Hospital and University Clinics, Auenbruggerplatz 2, 8036 Graz, Austria, Europe; CBmed GmbH — Center for Biomarker Research in Medicine; E-mail: markus.kreuzthaler@medunigraz.at.

The idea that character-level neural language models (CNLM) alone, based on recurrent neural networks (RNN) and more specifically long short-term memories (LSTM) are able to sufficiently capture the information needed for a given task recently got again attention by various research groups. The main motivation behind these experiments — Google named them “tokenizer-free language models” [1] — is to address NLP tasks using only single characters as input representations, thus capturing human language on its most granular representation level [2].

In this work we investigated the possibility of character-based *clinical* language modelling in order to address its well-known idiosyncrasies (telegram-style, single-word composition, spelling variants, and typing errors) with the purpose of optimally capturing the meaning of real-world textual data.

The paper is structured as follows: Section 2 gives an overview of recent work on CNLMs. Section 3 describes the clinical data set for the investigation as well as used methods and language models. Section 4 exemplifies the results, gives insight into the lexical semantics of the obtained embeddings and discusses the results. Finally, Section 5 concludes the paper and provides an outlook for further investigation.

2. Background and Related Work

Zhang and LeCun [3] applied convolutional neural networks (CNN) on a single character input level for ontology classification, sentiment analysis and text categorization. They stacked six convolutional layers and three fully-connected layers for the specific tasks. The experimental design reached a test set based accuracy of 0.98 for ontology classification (DBpedia), 0.95 for sentiment analysis using the Amazon review polarity data set, and 0.60 exploiting the Amazon review full score data set. An accuracy of 0.71 was reached for the Yahoo! Answers topic classification data set.

Kim et al. [4] applied single character-level inputs to a deep neural network model consisting of a CNN, a highway network over characters, and an LSTM at the end. They tested their model on the Penn Treebank for different languages and achieved state-of-the-art results with less parameters in comparison to the other machine learning based approaches. They reported a perplexity of 78.9 on the English Penn Treebank test set. Their character-level input models outperformed word/morpheme-based input representation schemes in general. The model was also shown to be able to capture different levels of semantics on a token level constituted by a time-series based input representation scheme of single characters.

More recently, Choe et al. [1] from Google showed that tokenizer-free CNLMs are competitive with current state-of-the-art language models using the One Billion Word benchmark in combination with a deep transformer neural network containing 40 self-attention layers. By using several semantic benchmarks, they showed the ability of their network to capture layer-wise word-level semantics even though the input representation was character-based. This work corroborates another recent work from Stanford and Facebook [5], which concluded that CNLMs based on recurrent neural networks (e.g. LSTMs) are able to capture upper-level semantics.

Motivated by the research mentioned above, we explored to what degree CNLMs can capture token-level lexical semantics from clinical texts. We accomplished that by fetching embeddings out of an RNN, after training it on a critical mass of real-world data,

viz. ICD-10 coded problem list entries from a clinical information system, modelled as a time series of one-hot encoded single characters. With a nearest neighbour search on selected clinical narrative content we fetched word types within the embedding space and hypothesized that they were semantically connected to the given input. To the best of our knowledge, this is the first investigation on the extent of CNLM-based lexical semantics that can be obtained from clinical narratives.

3. Methods and Data

The initial data set is composed of ~ 1.9 million unique 50 characters long problem list descriptions in German, annotated with ICD codes. These annotations had been done by physicians on in-patient discharge, the ICD codes being relevant for billing. To have a sufficient number of initial samples per class we took the top-100 most frequent three-character ICD-10 codes from a 90% split and up-sampled them to the most common one to balance the whole data set. The final data consisted of ~ 6 million coded problem list entries used for CNLM training downstreamed to their corresponding disease code.

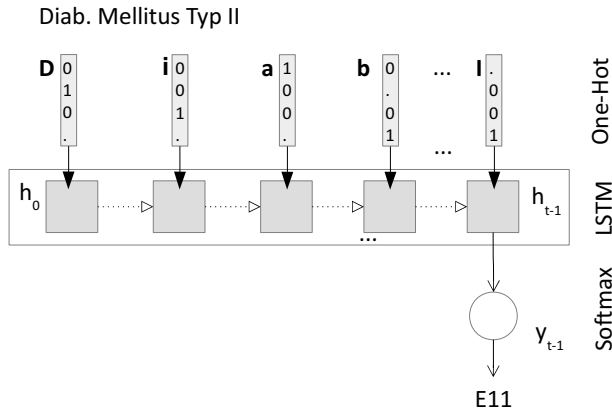


Figure 1. Schematic description of the LSTM network in use [6].

After having prepared the training data set, we fed it to an LSTM network as depicted in Figure 1. The problem list entry is represented as a series of one-hot encoded input vectors of dimensionality 122 (number of alphabetic characters in the training corpus) to the recurrent neural network. We refrained from any pre-processing and used all strings as they were in the raw data. The neural network was trained for 5 epochs (~ 73 hours) using the `deeplearning4j`² framework on an NVIDIA Titan V GPU with the following parameters: mini-batch size: 512; learning rate: 1.0; L2 regularisation: $1E-6$; weight initialisation: XAVIER; updater: RmsProp; LSTM hidden state vector size: 128; number of LSTM cells: 60; LSTM activation: TANH; output layer activation: SOFTMAX; output layer loss function: MCXENT; total number of network parameters: 141,412, which is about a factor of 40 less than available training observations.

²<https://deeplearning4j.org/>

In order to obtain a specific character-induced word embedding, we accessed the hidden state h_{t-1} out of the cell at step $t - 1$ for a given token \mathbf{X} of length t . Using the k-d tree data structure we built a searchable $k = 128$ dimensional index for all occurring types in the training corpus to get back the *top n* nearest neighbours given an input vector. The input vector was inferred the same way using the trained LSTM network as described before. With the trained LSTM network and the k-d tree data structure, we could obtain the nearest neighbours in reasonable time by exploiting the cosine similarity measure in the embedding space. This space was used to test our hypothesis that nearest neighbours to a given input vector contain verifiable semantics with respect to the initial search vector.

4. Results and Discussion

Table 1 lists the top 10 nearest neighbours to a given concept in descending order. The first column contains the results for the disorder “Nikotinabusus” (nicotine abuse). Interestingly, most of the nearest embeddings to the given token contain variations of the clinical quantification of smoking cigarettes in pack years (py), a level of detail that is irrelevant for ICD. The pack year (py) is the unit in which the inhaled smoke dose of a cigarette smoker is approximately described. Moreover, the quantitative values correspond to the manifestation of a harmful use of nicotine via some smoking behaviour.

“Nikotinabusus”		“Handverletzung”	
Word type	Sim.	Word type	Sim.
chronisch	1.00	Handverletzung	1.00
(15/die)	0.99	Strecksehnendurchtr	0.91
abuse	0.99	Fingernagelbruch	0.81
(20Stk./d)	0.99	Zeigefingergrundgliedfraktur	0.80
(25/d)	0.99	ausw.kons.	0.80
Z/d)	0.99	Handwurzelkontusion	0.79
(135	0.99	eingestauchte	0.79
(Abstin.seit	0.98	Daumenendphalanxstauchungsfraktur	0.78
(30pys)	0.98	Handwurzelknochen-Fraktur	0.78
(50pj)	0.98	4.links	0.77

Table 1. Top 10 nearest neighbors with corresponding similarity scores for two example words.

The second column contains the most similar character-level-induced word embeddings for “Handverletzung” (hand injury). The majority of the top 10 results contain different types of injuries of the body structure hand, and anatomical locations which are part of the hand. A remarkable result is the semantic response “Daumenendphalanxstauchungsfraktur” (thumb end phalanx compression fracture). Semantically obviously related to each other, the lexical representation of “Handverletzung” and “Daumenendphalanxstauchungsfraktur” has nearly nothing in common with a Levenshtein distance of 25.

5. Conclusion and Outlook

In this paper we investigated to what extent a CNLM with an LSTM network architecture is able to induce meaningful lexical semantics in the clinical domain. We trained the network on ~ 6 million ICD-10 coded problem lists, a critical amount of training data for language modelling. After the training, we fetched the embeddings by accessing the hidden state h_{t-1} within the cell at the last step of an input word of the LSTM network. We then indexed the resulting 128-dimensional character-level-induced word embeddings using a k-d tree data structure and, via a top 10 nearest neighbour search, obtained the most similar whitespace separated tokens given an input word. We showed that meaningful lexical semantics could be captured at this representation level using an exploratory evaluation.

Our results support recent investigations that single character-level representations to RNNs are able to fetch semantics in contrast to the most low-level fine-granular input scheme. The work presented here uses a single LSTM network in contrast to more complex models from recent research on deep transformer models like BERT [7] or stacked Bi-LSTMs like ELMo [8] for contextual embedding modelling.

Furthermore, this preliminary evaluation was done purely on a human interpretation level on selected examples. Future investigations of CNLMs in the clinical domain should consider objective evaluation measurements for semantic similarity and language model quality. Besides the mentioned limitations, the work presented in this paper supports the promising potential of CNLMs in processing raw clinical language in a robust but nevertheless semantically precise manner.

References

- [1] Dokook Choe, Rami Al-Rfou, Mandy Guo, Heeyoung Lee, and Noah Constant. Bridging the gap for tokenizer-free language models. *arXiv preprint arXiv:1908.10322*, 2019.
- [2] Alan Akbik, Duncan Blythe, and Roland Vollgraf. Contextual string embeddings for sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649, 2018.
- [3] Xiang Zhang and Yann LeCun. Text understanding from scratch. *arXiv preprint arXiv:1502.01710*, 2015.
- [4] Yoon Kim, Yacine Jernite, David Sontag, and Alexander M Rush. Character-aware neural language models. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [5] Michael Hahn and Marco Baroni. Tabula nearly rasa: Probing the linguistic knowledge of character-level neural language models trained on unsegmented text. *arXiv preprint arXiv:1906.07285*, 2019.
- [6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [7] Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*, 2019.
- [8] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.

Clinical Abbreviation Disambiguation Using Deep Contextualized Representation

Mingkai PENG^{a,1} and Hude QUAN^a

^a*Department of Community Health Sciences, University of Calgary, Calgary, Canada*

Abstract. The objective of this study is to develop a method for clinical abbreviation disambiguation using deep contextualized representation and cluster analysis. We employed the pre-trained BioELMo language model to generate the contextualized word vector for abbreviations within each instance. Then principal component analysis was conducted on word vectors to reduce the dimension. K-Means cluster analysis was conducted for each abbreviation and the sense for a cluster was assigned based on the majority vote of annotations. Our method achieved an average accuracy of around 95% in 74 abbreviations. Simulation showed that each cluster required the annotation of 5 samples to determine its sense.

Keywords. clinical abbreviations, deep contextualized representation, cluster analysis

1. Introduction

Abbreviations are ubiquitous in clinical notes as they are a convenient way for health care providers to record clinical concepts such as diseases and procedures within limited time. However, ambiguity of clinical abbreviations is very common. It was reported that 33.1% of abbreviations in the Unified Medical Language Systems (UMLS) have more than one sense[1]. Abbreviation disambiguation is the process of identifying the correct sense of an abbreviation within a given context and is an important step to automate information extraction from clinical notes. Understanding clinical abbreviations still remains a challenging task for current clinical natural language processing (NLP) systems due to two major reasons: lack of a complete sense inventory of abbreviations, and efficient methods to disambiguate abbreviations with multiple senses[2].

Overall, supervised learning performs well in clinical abbreviation disambiguation[3]. These methods usually build a classifier for each abbreviation using an annotated corpus with a large number of instances. However, the annotation process to create training data is expensive and time-consuming, which limits the scalability of supervised learning approaches. Many studies have explored different ways to reduce the annotation efforts for abbreviation disambiguation. Semi-supervised approaches have been explored to generate training data by submitting the long forms with their corresponding abbreviations and then training a classifier on the generated data[3]. The semi-supervised approach can achieve up to 90% of classification accuracy. To build the sense inventory, instances of an abbreviation were clustered into different sense clusters

¹ Corresponding Author, 2500 University Drive NW, University of Calgary, Calgary, Canada; E-mail: mpeng@ucalgary.ca.

and selected instances from each cluster were manually annotated to create the final list of senses[4]. The clustering-based method significantly reduced the annotation cost and increased the sense inventories.

Previous studies have demonstrated the usefulness of word embedding for abbreviation disambiguation. Pooled neighbor word embedding of the abbreviations were used to train support vector machine classifiers for each abbreviation and the classifiers achieved good results on two clinical abbreviation datasets[5]. Use of deep contextualized word representation from pre-trained language model on large corpus has achieved state of the art performance on many NLP tasks, including abbreviation disambiguation[6][7].

In this study, we explored the use of deep contextualized word representations as the input features and the cluster analysis to divide instances into clusters for annotation. The annotated cluster would be used for subsequent classification.

2. Data and Methods

In this study we used the clinical abbreviation sense inventories from University of Minnesota (UMN)[8]. The inventory was derived from 352,267 clinical notes including admission notes, consultation notes and discharge summaries and included 74 abbreviations. For each abbreviation, 500 instances were randomly selected and annotated. For each instance, 10 words pre and post the abbreviations were extracted and used in this study.

The pre-trained BioELMo was used to get the contextualized representation of an abbreviation within each instance. Embedding from Language Models (ELMo) is a deep bidirectional long short-term memory (LSTM) language model with two hidden layers and achieves the state of the art results in several NLP problems (e.g. word sense disambiguation) due to its deep contextualized representation[6]. Embedding representation of each token in ELMo is a function of the entire input sentence. We used the BioELMo that was trained on the PubMed corpus with 10 million abstracts[9]. The first and second layers from BioELMo and summation of the two layers were used as the feature presentation in our study.

Instances from each abbreviation were divided into training (400) and test (100) data. Principal component analysis (PCA) was conducted on the word vectors of training data for dimension reduction. The PCA retained 95% of the total variance. The K-Means cluster algorithm was applied to extract 10, 15, and 20 clusters respectively. The senses of clusters were determined based on the majority vote of its cluster members. The trained PCA and cluster models were applied to test data to classify its sense. Simulation was conducted to evaluate the number of samples to be reviewed to determine the sense for a cluster. After the cluster analysis on the training data, we randomly sampled 1, 3 or 5 instances for each cluster per abbreviation to determine the sense of the cluster based on the majority vote. See Figure 1 for an overview of preprocessing and analysis process. The accuracy of classification was used to assess the performance of method.

3. Results

There were 74 abbreviations in the clinical sense inventory. Half of them had 3 senses and 8 abbreviations had more than 4 senses (Figure 2). Most of abbreviations had

unbalanced distribution of senses. 68 abbreviations had a sense accounting for over 50% of instances and 35 abbreviations had a dominant sense for more than 80% of instances.

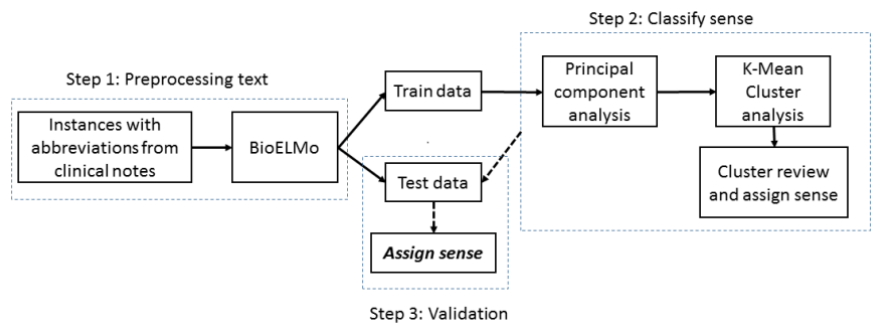


Figure 1. An overview of the method for clinical sense disambiguation.

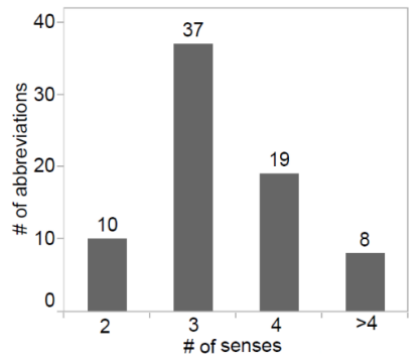


Figure 2. Distribution of senses for the abbreviations.

Table 1. Accuracy of the method using different bioELMo layers and the number of clusters

Layer	Number of cluster	Accuracy on training data	Accuracy on test data
Layer 1	10	0.912 (0.072)	0.916 (0.087)
Layer 1	15	0.921 (0.065)	0.924 (0.076)
Layer 1	20	0.927 (0.064)	0.929 (0.076)
Layer 2	10	0.936 (0.060)	0.943 (0.065)
Layer 2	15	0.943 (0.057)	0.949 (0.062)
Layer 2	20	0.946 (0.051)	0.954 (0.059)
Sum of layer 1 and 2	10	0.930 (0.066)	0.935 (0.070)
Sum of layer 1 and 2	15	0.938 (0.056)	0.944 (0.058)
Sum of layer 1 and 2	20	0.943 (0.053)	0.948 (0.058)

Impact of ELMo layer and number of cluster on classification accuracy are presented in Table 1. Overall, use of layer 2 achieved better accuracy than use of layer 1 or sum of layer 1 and 2, and 20 clusters in the cluster analysis achieved the highest overall accuracy. Due to the nature of unsupervised learning used in our study, the differences in performance between training and test data were minimal.

Detailed results for each abbreviation are shown in Figure 3. For abbreviation with only 2 sense, there were minimal increases in accuracy when the number of clusters increased from 10 to 20. However, increasing the number of clusters improved the

classification accuracy in abbreviation with multiple senses. There were slight decrease of accuracy with the increase of senses, in particular for abbreviations with > 4 senses.

Simulation was conducted to evaluate how many instances were needed to be reviewed to determine the sense of a cluster. As expected, the increase of number of reviewed instances improved the classification accuracy. Use of 5 instances to determine the sense of a cluster achieved comparable overall accuracy to the use of all 400 instances.

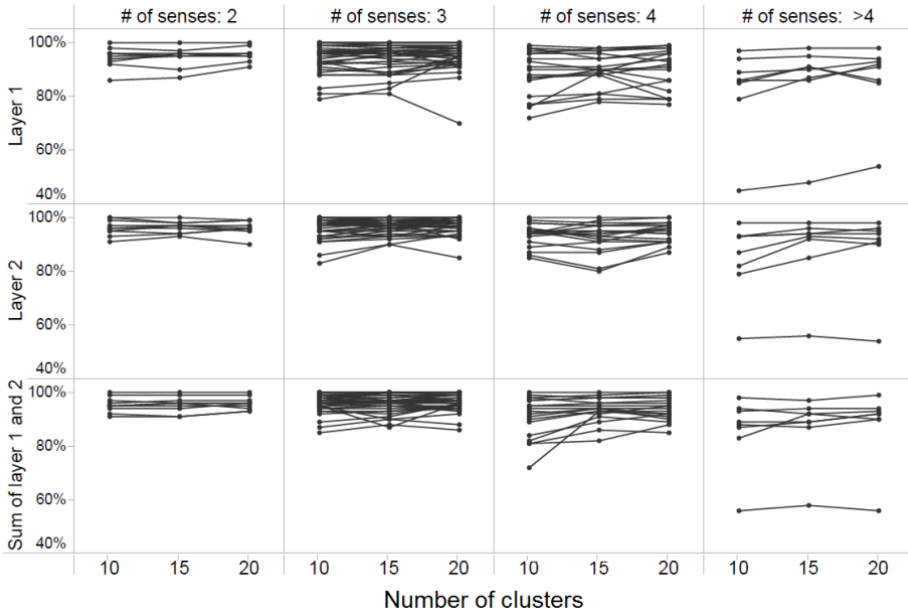


Figure 3. Impact of layer and number of cluster on the classification accuracy in test data. Each line represents one abbreviation.

4. Discussion

Our study showed that use of deep contextualized representations as features and cluster analysis to disambiguate clinical abbreviations achieved similar accuracy as supervised learning methods. Use of layer 2 of BioELMo and 20 clusters achieved the best performance. Simulation analysis demonstrated the potential cost-efficiency for sense disambiguation as it only required 5 samples per cluster without the loss of accuracy.

Use of word embeddings from pre-trained language models can capture rich semantic and contextual information for abbreviations and improve the accuracy of abbreviation disambiguation. The previous study used a bag of word representation, position information, and section name of clinical notes where the abbreviation occurs as the input features for cluster analysis and achieved an average accuracy of 93.7% on 12 abbreviations[4]. The use of deep contextualized representation as the input for a neural network model achieved an average accuracy of 98.1% on 11 clinical abbreviations[7]. This clearly shows the benefit of deep contextualized representation for abbreviation disambiguation.

Our method shows promising results on abbreviation disambiguation while requiring around 100 instances (20 clusters x 5 instances) for annotation to achieve comparable results to the supervised learning approaches. This could significantly reduce the cost of annotation. A recent study proposed the use of interactive learning algorithms to train abbreviation disambiguation classifiers using human created features[10]. During the annotation process, an expert was required to specify the sense of the indicated word and highlight the supporting evidence in a labeled instance. The algorithms achieved 90% accuracy using only 15 labels on the UMN dataset. Our method can achieve comparable results without the use of expert domain knowledge and steps to manually create the features for classification. This will speed up the development of new abbreviation disambiguation methods for new data sources.

Our study only applied the method in one corpus. Further investigation is required to test the generalizability of the method to other corpora.

5. Conclusion

We proposed the use of pre-trained language to extract deep contextualized representations for cluster analysis to disambiguate clinical abbreviations. The results show that with minimal annotation effort to determine the sense of each cluster, our method can achieve comparable accuracy as supervised learning methods.

References

- [1] H. Liu, Y.A. Lussier, and C. Friedman, A study of abbreviations in the UMLS, *Proc AMIA Symp.* (2001), 393–397..
- [2] Y. Wu, J.C. Denny, S. Trent Rosenbloom, R.A. Miller, D.A. Giuse, L. Wang, C. Blanquicett, E. Soysal, J. Xu, and H. Xu, A long journey to short abbreviations: Developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD), *J. Am. Med. Informatics Assoc.* 24 (2017), e79–e86.
- [3] G.P. Finley, S.V.S. Pakhomov, R. McEwan, and G.B. Melton, Towards Comprehensive Clinical Abbreviation Disambiguation Using Machine-Labeled Training Data. *Proc AMIA Symp.* (2016), 560–569.
- [4] H. Xu, P.D. Stetson, and C. Friedman, Methods for Building Sense Inventories of Abbreviations in Clinical Notes, *JAMA.* 16 (2009), 103–108.
- [5] Y. Wu, J. Xu, Y. Zhang, and H. Xu, Clinical Abbreviation Disambiguation Using Neural Word Embeddings, *ACL-IJCNLP* (2015), 171-176.
- [6] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, Deep Contextualized Word Representations, *NAACL* (2018), 2227–2237.
- [7] Q. Jin, J. Liu, and X. Lu, Deep Contextualized Biomedical Abbreviation Expansion, *ACL-IJCNLP*, (2019), 88–96.
- [8] S. Moon, S. Pakhomov, N. Liu, J.O. Ryan, and G.B. Melton, A sense inventory for clinical abbreviations and acronyms created using clinical notes and medical dictionary resources, *J. Am. Med. Informatics Assoc.* 21 (2014), 299–307.
- [9] Q. Jin, B. Dhingra, W. Cohen, and X. Lu, Probing Biomedical Embeddings from Language Models, *ArXiv* (2019), 1904.02181,
- [10] Y. Wang, K. Zheng, H. Xu, and Q. Mei, Interactive medical word sense disambiguation through informed learning, *J. Am. Med. Informatics Assoc.* 25 (2018), 800–808.

Clinical Concept Normalization on Medical Records Using Word Embeddings and Heuristics

João Figueira SILVA ^a, Rui ANTUNES ^a, João Rafael ALMEIDA ^{a,b}, and Sérgio MATOS ^{a,1}

^a DETI/IEETA, University of Aveiro, Portugal

^b Department of Computation, University of A Coruña, Spain

Abstract. Electronic health records contain valuable information on patients' clinical history in the form of free text. Manually analyzing millions of these documents is unfeasible and automatic natural language processing methods are essential for efficiently exploiting these data. Within this, normalization of clinical entities, where the aim is to link entity mentions to reference vocabularies, is of utmost importance to successfully extract knowledge from clinical narratives. In this paper we present sieve-based models combined with heuristics and word embeddings and present results of our participation in the 2019 n2c2 (National NLP Clinical Challenges) shared-task on clinical concept normalization.

Keywords. natural language processing, clinical information extraction, clinical concept disambiguation, word embeddings, sieve-based model

1. Introduction

Electronic health records (EHRs) hold great value for the correct understanding of patient trajectories since they combine the multimodality in medical data with the temporal aspect that is embedded in the clinical history, rendering it as a vital component to correctly understand how symptoms, findings, diagnoses and diseases evolve [1].

Textual data is highly relevant and can be found in EHRs in (semi-)structured and unstructured formats, the latter being commonly referred to as free text. Medical narratives (e.g. discharge or admission reports) are stored as free text since natural language provides a flexible convoy for physicians to track and report each medical situation, overcoming the limitations from ambiguous and unspecific terms that physicians can find when using coding standards such as RxNorm or SNOMED-CT. Owing to that, it is reckoned that free text frequently contains plenty of information otherwise not obtainable from other data sources [2]. However, it is unfeasible to manually analyze large scale medical datasets. Despite the intricacies inherent to free text that make it very challenging to process and interpret, the process of automatically annotating clinical narratives is

¹Corresponding Author: Sérgio Matos, DETI/IEETA, Universidade de Aveiro, Campus Universitário do Santiago, Aveiro 3810-193, Portugal; E-mail: aleixomatos@ua.pt.

an important way to summarize or extract relevant data from medical text. This typically requires the text processing tasks of named entity recognition (NER) and named entity normalization (NEN) to extract relevant concepts and to standardize their referencing. The latter is a necessary step to combat ambiguities since clinical text contains many abbreviations, misspellings, and domain-specific expressions.

In this paper we describe an approach for normalization of clinical entity mentions using sieve-based models combined with heuristics and word embeddings. The proposed method was used in the 2019 n2c2 shared task on clinical concept normalization³, where the aim was to link clinical entities to SNOMED CT or RxNorm vocabularies through UMLS Concept Unique Identifiers (CUIs) [3].

2. Materials and Methods

The aim of the n2c2 clinical concept normalization task was to normalize medical entities to standard medical vocabularies. For simplicity purposes, the task focused only in NEN, bypassing the NER step by providing relevant clinical mentions identified *a priori*.

In this section, we describe the dataset used as well as the three different methodologies applied, namely a method based on concept embeddings, an improved version of a baseline sieve approach, and a final rule-based method.

2.1 Dataset

The Medical Concept Normalization (MCN) corpus proposed by Luo *et al.* [1] was used. It comprises a wider set of clinical concepts in contrast to previous NEN challenges that only evaluated the normalization of disease mentions [4,5].

Each annotated clinical entity is associated with a single CUI from the UMLS 2017AB version. For example, “hypertension” and “HTN”, or “blood pressure” and “BP” are two examples of expressions that refer to the same concepts and are therefore identified by the same CUIs (C0020538 and C0005824, respectively). Although UMLS encompasses several vocabularies, only two were used for annotation. RxNorm was used to annotate clinical drugs and medications, whereas SNOMED-CT, an extensive vocabulary of clinical terminology, was used for normalizing the remaining concepts (disorders, procedures, body structures, and others).

Table 1. Detailed dataset statistics.

	Training	Test	Total
Number of clinical records	50	50	100
Number of annotated entities	6 684	6 925	13 609
Number of unique CUIs	2 331	2 579	3 792

The dataset contains a total of 100 annotated discharge summaries and is split in training and test subsets (Table 1). This allowed model development in the training set and a blind official challenge evaluation using withheld test data (the gold standard annotations for the test data were only available after the official evaluation).

2.2 Developed Methodologies

2.2.1 Concept Embeddings Similarity

This method involves two sequential steps: (1) text pre-processing and (2) similarity computation. In the first step, certain text rewrite rules were handcrafted by inspecting the clinical named entities in the training set (Table 2) with the aim of cleansing the surface representation of these mentions. In addition to the text replacements made, HTML entities and other superfluous symbols were also discarded.

Table 2. Examples of text rewrite rules handcrafted according to the training set.

b/l	bilateral	po	oral	10%	partial
co2	carbon dioxide	trop	troponin	¼	fourth
e.	escherichia	u/s	ultrasound scan	x2	double
iv	intravenous	vit	vitamin	2L	two liters
mso4	morphine sulphate	w/u	workup	3 of 6	iii/vi

In the second step we represented the clinical named entities (from training and test sets) and UMLS concept names by using pre-calculated biomedical word embeddings. We employed the publicly available BioWordVec model [6] that was created applying the fastText library [7] and was generated from over 30 million documents from PubMed articles and clinical notes from the MIMIC-III database.

From these representations we created a direct mapping between terms and CUIs where each term was defined by the embedding vector average of its constituent words [8]. This mapping was built using (i) text mentions from the train set and (ii) concept names from the UMLS CUIs corresponding to the SNOMED-CT and RxNorm vocabularies.

At last, for making the predictions in the test set, the cosine similarity was calculated between each test mention embedding vector and all pre-calculated term embeddings [8]. The CUI corresponding to the most similar term was chosen.

A simplified overview of this method is shown in Figure 1. Due to its ability to correctly capture text mentions with similar semantics, this method was also used as a final sieve to predict remaining (unclassified) entities in the next two pipelines.

2.2.2 Improved Sieve-Based Approach

The sieve-based approach by Luo *et al.* [1] was firstly deployed, containing exact matching with the training set annotations and with UMLS, along with MetaMap sieves [9]. This pipeline was performed in two stages, firstly using raw text mentions and secondly using clean text mentions. This approach yielded a 5-fold accuracy on the training set of 0.783. An error analysis showed that MetaMap generated too many incorrect identifiers, limiting the number of unclassified mentions passing to the second sieve stage, thus capping the maximum attainable accuracy.

The pipeline was reworked by shifting the MetaMap sieve from stage 1 to the end of the pipeline, right before the MetaMap sieve from stage 2. Further improvements involved performing a Monte Carlo simulation to select an optimal threshold for CUI classification with MetaMap, greatly reducing false positives. Remaining unclassified mentions were classified using the concept embeddings approach described in Section 2.2.1. The final pipeline is presented in Figure 1.

2.2.3 An Orchestrator for Combining Simple and Complex Rules

This approach applies a 4-step pipeline combining rules and dictionary lookup, followed by previously computed concept embeddings.

The initial step processes mentions containing ambiguous terms that map to multiple CUIs in the training set. For each mention, several rules are applied that consider specific words from the respective sentence and section heading if applicable. In steps 2 and 3, the system performs an exact match using dictionaries created from the training set and the UMLS database, respectively. This three step pipeline is invoked twice: in a first phase it is applied to raw mentions, which was essential to detect certain patterns that could match concepts annotated in the training set; in a second phase, the remaining mentions are pre-processed as described in Section 2.2.1 before the pipeline is applied.

Finally, we combined this workflow with the concept embeddings from Section 2.2.1 by having a final stage where concepts that completed the previous workflow without an attributed CUI were processed using the concept embeddings.

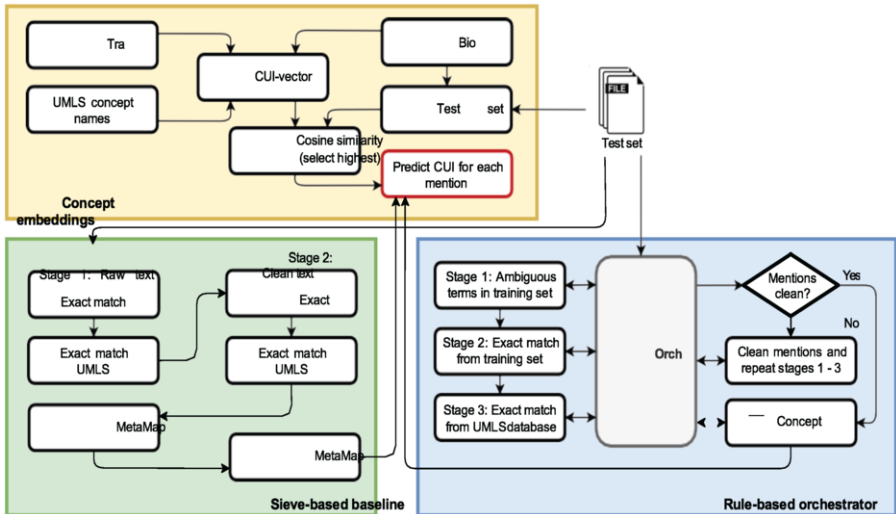


Figure 1. Final system architecture composed of three different approaches.

3. Results and Discussion

Table 3 shows the results, obtained in the training and test sets, of the three previously described methodologies.

Concept embeddings achieved an accuracy of 0.812 and 0.801 in the training and test sets. However, a posterior evaluation without using handcrafted replacements proved that the created text patterns were biased into the training set and led to overfitting, since simpler text pre-processing resulted in a lower training accuracy (0.807) and similar test accuracy (0.800).

The rule-based approach combined with the concept embeddings method obtained the highest accuracy in the test set with a value of 0.806, corresponding to a 4 percentage points improvement compared to the baseline sieve-based model proposed by Luo et al. [1] that obtained an accuracy of 0.764 in the test set.

Table 3. Obtained accuracy results with the three distinct methodologies. Results in the training set achieved by 5-fold cross-validation (10 repetitions in the concept embeddings method).

	Concept embeddings	Sieve-based	Rule-based
Training	0.812	0.806	—
Test	0.801	0.791	0.806

4. Conclusions and Future Work

In this paper, we proposed three methodologies to normalize clinical concepts from patient clinical reports automatically. The development of such methods applied in health-care data migration pipelines is essential to increase data value and have harmonized datasets.

As future work, we intend to improve each methodology individually by using information from the surrounding context and section heading of each mention. Also we intend to incorporate UMLS concept definitions to enrich the semantic knowledge regarding each concept. These improvements may reduce concept ambiguity issues. As a final refinement, we aim to apply deep neural network models to predict the correct CUI embedding vector.

5. Acknowledgments

This work has received support from the EU/EFPIA Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 806968. João Figueira Silva, Rui Antunes, and João Rafael Almeida are funded by the FCT - Foundation for Science and Technology (national funds) under the grants PD/BD/142878/2018, SFRH/BD/137000/2018, and SFRH/BD/147837/2019 respectively.

References

- [1] Yen-Fu Luo, Weiyi Sun, and Anna Rumshisky. MCN: a comprehensive corpus for medical concept normalization. *Journal of Biomedical Informatics*, 92:103132, April 2019.
- [2] Kasper Jensen, Cristina Soguero-Ruiz, et al. Analysis of free text in electronic health records for identification of cancer patient trajectories. *Scientific Reports*, 7(46226), April 2017.
- [3] Olivier Bodenreider. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(Suppl 1):D267, January 2004.
- [4] Sameer Pradhan, Noemie Elhadad, et al. Task 1: ShARe/CLEF eHealth Evaluation Lab 2013. In *CLEF (Working Notes)*, Valencia, Spain, September 2013. CEUR Workshop Proceedings.
- [5] Sameer Pradhan, Noémie Elhadad, et al. SemEval-2014 Task 7: analysis of clinical text. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 54–62, Dublin, Ireland, August 2014. Association for Computational Linguistics.
- [6] Qingyu Chen, Yifan Peng, and Zhiyong Lu. BioSentVec: creating sentence embeddings for biomedical texts. *arXiv:1810.09302*, October 2018.
- [7] Piotr Bojanowski, Edouard Grave, et al. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5(1):135–146, 2017.
- [8] Tomas Mikolov, Kai Chen, et al. Efficient estimation of word representations in vector space.
- [9] *arXiv:1301.3781*, January 2013.
- [10] Alan R. Aronson. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In *Proceedings of the AMLA Symposium*, pages 17–21, Washington, DC, USA, November 2001. American Medical Informatics Association.

Clinical History Segment Extraction from Chronic Fatigue Syndrome Assessments to Model Disease Trajectories

Sonia PRIOU^a, Natalia VIANI^a, Veshalee VERNUGOPAN^b, Chloe TYTHERLEIGH^a, Faduma Abdalla HASSAN^c, Rina DUTTA^a, Trudie CHALDER^a and Sumithra VELUPILLAI^{a,1}

^aIoPPN, King's College London; NIHR Maudsley Biomedical Research Centre

^bUniversity of Glasgow, School of Medicine

^cLeicester Medical School

Abstract. Chronic fatigue syndrome (CFS) is a long-term illness with a wide range of symptoms and condition trajectories. To improve the understanding of these, automated analysis of large amounts of patient data holds promise. Routinely documented assessments are useful for large-scale analysis, however relevant information is mainly in free text. As a first step to extract symptom and condition trajectories, natural language processing (NLP) methods are useful to identify important textual content and relevant information. In this paper, we propose an agnostic NLP method of extracting segments of patients' clinical histories in CFS assessments. Moreover, we present initial results on the advantage of using these segments to quantify and analyse the presence of certain clinically relevant concepts.

Keywords. Natural Language Processing, Chronic Fatigue Syndrome; Electronic Health Records; Clinical Informatics

1. Introduction

Chronic fatigue syndrome (CFS) is a debilitating illness characterised by extreme tiredness, with a wide range of related symptoms, including muscle fatigue and sleep disturbance [1]. Given that the origin of the condition is still unknown, analysing large amounts of patient data is an important step towards an improved understanding of its aetiology and clinical trajectories. Clinical assessments of CFS provide a comprehensive overview of patients' symptoms and functional limitations, and are usually conducted when a patient is referred to a specialist clinic. These assessments include past history as well as current situation, thus representing a valuable source of data for large-scale analysis. In electronic health records (EHRs), however, clinical assessments are often primarily documented in an unstructured form (free text), which cannot be easily analysed on a large scale. To allow the automatic analysis of such texts, natural language processing (NLP) methods are becoming increasingly popular [2].

During a CFS assessment, an extensive amount of patient information (e.g., demographic data, medication history, treatment plans) is gathered. To automatically

¹ Corresponding Author; E-mail: sumithra.velupillai@kcl.ac.uk.

process these different input texts, capturing only information specifically on factors that may have contributed to or are associated with symptom and condition trajectories is helpful. For downstream tasks such as clinical concept extraction and trajectory modelling, NLP methods that identify useful segments within the text are needed, and these are usually found in paragraphs related to *clinical history*. Automatically extracting these relevant segments for large patient cohorts is a challenging task, especially due to the variability of EHR document structures across healthcare institutions. In initial clinical assessments, most clinicians document the presenting (or primary) complaint first, and then less strictly ordered are the history of present illness, past medical and psychiatric history, family history, social history, personal history and assessment & management plan [3]. However, these sections are not necessarily documented under specific headings. There have been several efforts to develop NLP approaches that automatically extract document sections from various types of clinical documents, e.g., discharge summaries and admission reports [3–5]; family history [6–8], or patient histories [9]. However, most of these studies rely on using section and subsection *heading* information, which is often not available in a predefined way.

As a crucial initial step to modelling symptom and condition trajectory patterns in CFS patients, extracting clinical history segments from initial assessments - regardless of the structure of the documents - is needed. In this work, we develop an agnostic method of extracting temporal segments of text that convey clinical history information. The method is evaluated against a small set of manually annotated CFS assessments. We also perform an initial experiment of the potential benefit of using these segments to analyse the presence of certain clinically relevant concepts in a larger CFS patient cohort.

2. Materials and Methods

This study is based on mental health records from the Clinical Record Interactive Search database [10] from the South London and Maudsley NHS Foundation Trust (Ethical approval: Oxford REC C, reference 18/SC/0372). The first clinical assessment for CFS patients usually consists of a comprehensive letter written by the clinician, including information on past and recent history. To develop and evaluate our segment extraction approach, we used “first” assessment letters for all patients diagnosed with CFS - based on the date, length (character count) and title of the documents. 1201 patients (one document per patient) were identified. Only 49.7% of the documents had some form of section headings. 36 CFS assessments were randomly extracted and manually annotated by two annotators (medical students, native English speakers). Annotators were asked to mark *sentences* that indicated history information (clinical, family, personal and social history). Here, we only focus on sentences marked as *clinical history*. This annotated dataset was randomly split into a training (20 docs) and blind test (16 docs) set.

For our segment extraction algorithm, we hypothesised that historical segments contain more time expressions (e.g. *in 2012*, *two years ago*) than other parts of the document. Time expressions were obtained using SUTime [11]. The segment extraction algorithm divides each document in sentences using a regular expression sentence tokenizer. Then, it flags each sentence as 1 if the sentence contains a time expression (date or duration), 0 otherwise. The first and last sentence of each document are always considered as unflagged (0) because of document headings and end signatures. To generate larger segments, additional sentences are then flagged 1 if the length of the segment flagged as 0 between two sentences marked 1 is shorter than the average length

of segments with sentences flagged as 0. Finally, flagged segments that are too short are deleted, based on calculating the average length of flagged segments.

Inter-annotation agreement (IAA) was calculated with f-measure on the sentence level, which approaches the κ statistic when the number of negative cases is large and unknown [12]. For the algorithm performance, recall (sensitivity) was the main quantitative evaluation metric, calculated at a sentence level as compared to the annotators (A1 and A2). To assess whether the extracted segments retained clinically relevant information, we searched for a set of pre-defined concepts relevant to CFS disease trajectories within these segments and in the entire text (using exact string match). We considered the following concepts, known to be related to CFS: stress, virus, depression, anxiety and suicidal thoughts/ideation. The number of concepts were counted and compared in extracted segments vs. entire documents. In addition, two clinicians analysed the qualitative nature of the extracted segments for 10 randomly extracted CFS documents, assessing whether they retained the key clinical information. False positives and false negatives from the temporal segmentation algorithm were also analysed, to gain more insight into how temporal information is documented in CFS assessments.

3. Results

IAA on the CFS dataset was 70.6 and 74.9 f-measure on the training and test set, respectively. Overall, the temporal segmentation algorithm resulted in an average recall of 72.3% on the training set. Results on the final test set were: 64.1%.

Figure 1 is a visualisation of a random subset of the segments (training set) found by the temporal segmentation algorithm (black) compared to the manually annotated sentences (pink). Each document is divided into sentences. When a sentence

(x-axis) has been marked as clinical history either by the algorithm (black) or the annotator (pink), it is marked as 1, 0 otherwise (y-axis). As can be observed, the algorithm tended to cover most annotations (i.e. high recall), but also captured sentences not annotated as clinical history (i.e. low precision). In the qualitative evaluation, the clinicians generally agreed that the extracted segments usually contained symptom and condition onset information, and that they represented clinical histories.

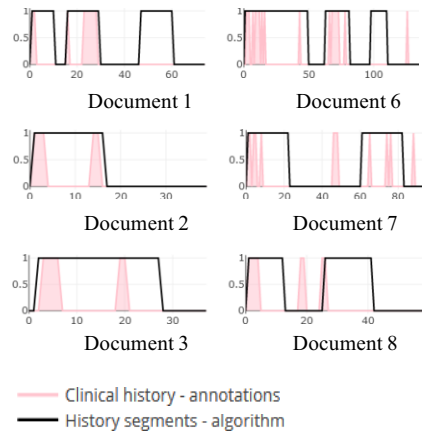


Figure 1: Visualisations of the temporal segmentation algorithm output as compared to manually annotated sentences (annotator A1) of clinical history on 6 documents from the CFS training set

Figure 2 shows results of the concept search on the entire set of CFS assessments (n=1,201). For 94% (1,133/1,201) of the patients (i.e. documents), at least one concept was found. ‘Stress’ was found in most assessments (69%, 832/1,201), and was, in the majority of cases, either in the extracted history segment (43%) or in both the history segment and outside (36%), i.e. in 21% the concept was documented only outside of the segment. ‘Virus’ was documented almost exclusively in the extracted history segments (11% were found outside of the segments). ‘Anxiety’ and ‘suicidal thoughts/ideation’ were more often reported outside of the history segments (30% and 34% respectively). ‘Suicidal thoughts/ideation’ was rarely found both within and outside the extracted segments (7%), and only found in 26% of the patients’ assessments (316). A manual analysis of 10 randomly extracted examples for ‘Anxiety’ and ‘Virus/viral’ (5 each) revealed some potential patterns. For ‘Anxiety’, all but one related either to speculations about the patient suffering from chronic anxiety disorder rather than chronic fatigue, or to anxiety related to work or health. For ‘Virus/viral’, 4/5 examples related to concepts that would be clinically relevant, and were all found at the beginning of the document, where the algorithm had failed to capture a segment. However, the segments in those documents contained synonym concepts such as ‘shingles’. The fifth example was a negation and was correctly excluded by the algorithm.

	Concepts found only in history segments	Some concepts found in history segments	No concepts found in history segments	Number of documents / patients
all concepts	26% [289]	64% [727]	10% [117]	1133
stress	43% [355]	36% [299]	21% [178]	832
anxiety	33% [181]	37% [199]	30% [163]	543
depression	48% [304]	28% [174]	24% [153]	631
virus / viral	73% [259]		16% [58]	356
suicidal thought(s) suicidal ideation	59% [185]	7% [24]	34% [107]	316

Figure 2: Coverage of pre-defined list of clinically relevant concepts in the entire set of CFS assessments

4. Discussion

To address the goal of finding population-level patterns in symptom and condition trajectories for CFS patients from their initial clinical assessments using NLP, selecting segments of the text describing the disease history is crucial. We have developed an algorithm that extracts temporal segments based on a simple approach relying on identifying time expressions (dates and durations) in sentences, and automatically expanding to segments by calculating patterns of consecutive sentences. As this method is agnostic to the data and does not make use of document section information, it can be extended to different types of clinical documents. In this approach, there is no assumption on the number of segments found per document. As shown in Figure 1, clinical history sentences can be either spread out across the document or gathered in one single segment. IAA agreement results were promising (within the ‘Substantial’ agreement range [13]), but further analysis and adjudication is needed. Most importantly, our current findings are based on a very small dataset – to generate more conclusive results, we will extend our study to larger datasets and perform additional experiments.

Previous studies on identifying segments in clinical narratives have heavily relied on the presence of pre-defined sections [3–5]. Only 50% of our data had section titles, thus such approaches would miss important information. We focused on improving recall, as our goal was to reduce the amount of irrelevant information in each document,

whilst ensuring we did not miss key clinical information. Generally, the algorithm covers the manual annotations (and more), but it also misses some sentences (Figure 1). False positives included examples of personal history rather than clinical history (#1 and #6) or current symptomology (#2). Several documents contained clinical history annotations at the end of the document (e.g. #6 and #7), that were missed by the current algorithm (false negatives). These corresponded to a brief overall summary which were repeating the information contained in previous segments. The algorithm also missed a few symptom-related sentences following a long social history paragraph (document #7). We plan to study alternative methods for automatically generating temporal segments based on larger datasets, as well as compare results with previously published approaches.

When searching for pre-defined clinically relevant CFS concepts, most documents had at least one (94%), and most concepts were found inside the extracted segments (90%). Concepts only found outside of segments were either mentioned in a different context than clinical history (e.g. the concept anxiety was used to describe the patient's current state) or used in the first sentence of the document (mainly for 'virus/viral'). We plan to extend this analysis by looking at additional concept mapping techniques such as word embeddings, clustering and topic modelling approaches.

Our main contributions in this study are: an agnostic method of extracting segments of EHR text that convey history information, and an initial experiment of the benefit of using these segments to analyse the presence of certain clinically relevant concepts in a CFS cohort. This is a first step towards large-scale studies on CFS disease trajectories.

References

- [1] A. Avellaneda Fernández, A. Pérez Martín, M. Izquierdo Martínez, et al., Chronic fatigue syndrome: aetiology, diagnosis and treatment, *BMC Psychiatry*. **9 Suppl 1** (2009) S1.
- [2] S. Sheikhalishahi, R. Miotto, J.T. Dudley, et al., Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review, *JMIR Med. Inform.* **7** (2019) e12239.
- [3] Y. Li, S. Lipsky Gorman, and N. Elhadad, Section Classification in Clinical Notes Using Supervised Hidden Markov Model, in: Proc. 1st ACM Int. Health Inform. Symp., ACM, New York, NY, USA, 2010: pp. 744–750.
- [4] J.C. Denny, A. Spickard, K.B. Johnson, et al., Evaluation of a method to identify and categorize section headers in clinical documents, *J. Am. Med. Inform. Assoc. JAMIA*. **16** (2009) 806–815.
- [5] P.J. Haug, X. Wu, J.P. Ferraro, et al., Developing a section labeler for clinical documents, *AMIA Annu. Symp. Proc. AMIA Symp.* **2014** (2014) 636–644.
- [6] R. Bill, S. Pakhomov, E.S. Chen, et al., Automated Extraction of Family History Information from Clinical Notes, *AMIA Annu. Symp. Proc.* **2014** (2014) 1709–1717.
- [7] J. Friedlin, and C.J. McDonald, Using a natural language processing system to extract and code family history data from admission reports, *AMIA Annu. Symp. Proc. AMIA Symp.* (2006) 925.
- [8] S. Goryachev, H. Kim, and Q. Zeng-Treitler, Identification and Extraction of Family History Information from Clinical Reports, *AMIA Annu. Symp. Proc.* **2008** (2008) 247–251.
- [9] W. Lee, and J. Choi, Temporal Segmentation for Capturing Snapshots of Patient Histories in Korean Clinical Narrative, *Healthc. Inform. Res.* **24** (2018) 179–186.
- [10] G. Perera, M. Broadbent, F. Callard, et al., Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource, *BMJ Open*. **6**(2016).
- [11] A.X. Chang, and C. Manning, SUTime: A library for recognizing and normalizing time expressions, in: Proc. Eighth Int. Conf. Lang. Resour. Eval. LREC-2012, European Languages Resources Association (ELRA), Istanbul, Turkey, 2012: pp. 3735–3740.
- [12] G. Hripcsak, and A.S. Rothschild, Agreement, the f-measure, and reliability in information retrieval, *J. Am. Med. Inform. Assoc.* **12** (2005) 296–298.
- [13] J.R. Landis, and G.G. Koch, The Measurement of Observer Agreement for Categorical Data, *Biometrics*. **33** (1977) 159.

Clinico-Environmental System for Personal Monitoring

Filipe PIRES^{1a,b}, André PEDROSA^{a,b}, João ALEGRIA^a and Carlos COSTA^{a,b},

^a *Department of Electronics, Telecommunications and Informatics, University of Aveiro*

^b *Institute of Electronics and Informatics Engineering of Aveiro*

Abstract. The ever-growing use of information and communication technologies in the past decades and the proliferation of mobile devices for monitoring vital signs and physical activity is enhancing the emergence of a new healthcare paradigm. More recently, citizens are becoming more sensible to the necessity of monitoring environmental health indicators and to its direct impact on personal health. This article proposes and describes the development of a clinico- environmental system for personal monitoring. The result is ContinuousCare, a personal healthcare information system that integrates personal smart devices with air quality monitors. The solution helps citizens to better understand their health and body activity with environmental context, aiding professional doctors with analysis tools and making available valuable data for external systems.

Keywords. clinico-environmental monitoring, health monitoring, personal health record, geolocation tracking

1. Introduction

The use of science and technology in medicine has been explored by mankind throughout the last decade more than ever. Information and communication technologies brought a proliferation of mobile devices for monitoring vital signs and physical activities of citizens and, consequently, the emergence of a new paradigm of medical monitoring is growing in the research communities.

On one hand, there already exists software to integrate distinct types of health monitoring devices and make available useful tools for injury detection and alert, amongst other purposes. Its target audiences are mostly specific groups like elderly people or athletes. On the other hand, the environmental conditions surrounding human life is often neglected by these healthcare technologies. Health disorders, painful conditions or even diseases can be directly correlated with local environmental conditions such as, for instance, high carbon dioxide levels. Current software solutions do not integrate environmental indicators, neglecting in consequence an important source of data than has tremendous potential to increase the prevention and diagnosis capabilities.

In this paper, we propose a personal monitoring solution called ContinuousCare that is able to integrate traditional personal health monitoring devices and environmental data sources, including home devices for air monitoring. Our system approaches the problem considering both the human activity and the environment it is exposed to and proposes powerful forms of data aggregation and visualization.

¹ Corresponding Author; E-mail: filipesnetopires@ua.pt

2. Background

A market research and literature reviewing process was performed for evaluation of actual proposal in the scope of ContinuousCare solution. It is possible to find a tremendous number of products, services and reports related to personal health monitoring, including geolocation recording. Due to the article size restrictions, we just report the ones that were more inspiring to the proposed system. An android application [1], developed by Turkish researchers, was the first product that takes advantage of geolocation to improve its service's quality. The application allowed users to keep track of elderly family members and be notified in case of emergencies, providing the closest hospitals. Health Tracking system [2] used GPS coordinates and other metrics to provide insights on the physical conditions of cyclists. The offered services seems very interesting and were inspiring for our proposal, being adapted to a more generalized target audience. Other solutions such as the Angel-Echo [3] or the UPHSM [4] provided us with perspectives on the potential of the subject and on the variety of use-cases and target audiences that could benefit from robust and intuitive health monitoring systems. Finally, the evaluation of key characteristics of already existing products, for similar problems, were taken in consideration during the design of ContinuousCare. A group of students from SSPACE college [5], in India, gave us valuable insights on what are the best metrics to take in consideration when monitoring body status.

3. System Architecture

The architecture of proposed personal monitoring system is shown in the Figure 1. The key design decisions aimed to ensure the robustness and scalability of ContinuousCare and its accessibility to any user with an internet connection. These regarded the collecting of data from our different sources, the appropriate storage strategy and the database load and access-time optimizations. It follows a modular structure where each module is replaceable and updates to one component do not imply the refactoring of others. The main component is the Monitoring Server that is divided into 3 functional units:

- Processing – it is responsible for gathering the information from supported devices and public data source APIs. This operation is scheduled, which means that it is executed many times per day. This unit is responsible for doing this scheduling and for processing the data collected (filtering it, assigning labels, and more);
- Persistence – it is responsible for storing all data after processing. It interacts with the database interface and ensures that every field is persisted correctly;
- REST API – it is the component responsible for supporting a REST (REpresentational State Transfer) interface for the system, serving as the entry point for every client application.

The second module is the GPS Service that was implemented since many personal monitoring devices do not provide access to its locations. Since geolocation is a crucial aspect in our proposal, it was necessary to implement from scratch a service responsible for receiving the users' locations and a mobile application that tracks them.

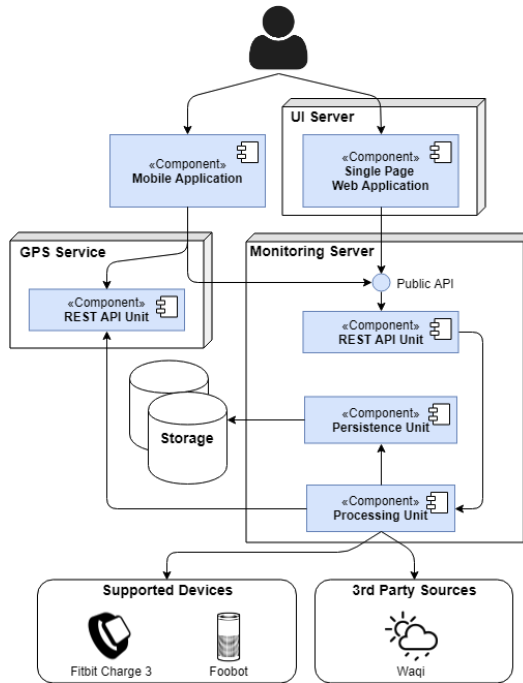


Figure 1. Architecture diagram.

In the persistence layer, we adopted a strategy based on a hybrid database. It was designed a relational database model in MySQL for the structured and interconnected data: accounts information, permissions between users, and devices. For the bulk data received from the devices and external sources, it was used a time series driven database, implemented with InfluxDB. This second one was found the best technology for our case as it is dedicated to continuous data streams and queries with time intervals. It is this database that feeds most of our web application's features. A system like ContinuousCare faced problems of big-data and some optimizations were performed, namely keeping in cache the values for a given time interval and periodically writing all the values stored with a batch process. In the long run, the database will hold a big load of data, some of it turning irrelevant as time passes. For this reason, the storage module supports the aggregation data after a defined period of time.

The platform provides a web application that runs on the UI Server and is our main interface with the end-users. It interacts with a public REST API (through the Monitoring Server) and makes use of the supported devices. It was developed with the NuxtJS framework, over VueJS, helping to maintain an intuitive website organization and giving us the possibility of deploying either a Single Page App or a Universal App (where pages are rendered on the server side for a smoother user experience). Several libraries for graphical elements, cookies and web-sockets were used, as well as an HTML template from Colorlib [6] for design purposes.

Finally, the system supports at this moment three types of data providers: 1) personal monitoring device – it was used the Fitbit Charge 3 [7], an advanced fitness tracker that has many health-related metrics. We focused on collecting calories, heart rate and sleep stages, as these present more interesting information for analysis

purposes; 2) air quality monitor – it was used Foobot [8], an indoor device able to detects invisible and odorless pollutants that fall into 3 categories (volatile compounds, particulate matter and carbon dioxide), calculates temperature and humidity levels and alerts users about unhealthy activities; 3) public portal with environment indicators – it was used the Wapi² public API that allows us to know the level of pollution of a given place on earth. By tracking the geolocation of our users, the system is able to decide whether to collect information from a Foobot or from Wapi according to the distance between the users and their respective air quality monitors.

4. Results

A Web platform was deployed on the Cloud with all functionalities described in this article. The system was continuously used by a user during one month for validation purposes. The platform supports two distinct user profiles, patients and doctors. Doctor accounts, once permitted by patients, have access to the profile information and to all records, grouped by Health Status, Environment, Sleep or Events. In each group, the user can see a particular metric in detail throughout time with the help of the line graphic. Users are able to add, remove or update monitoring devices. For instance, they can have many home devices in different places of common use (home, workplace, etc.).

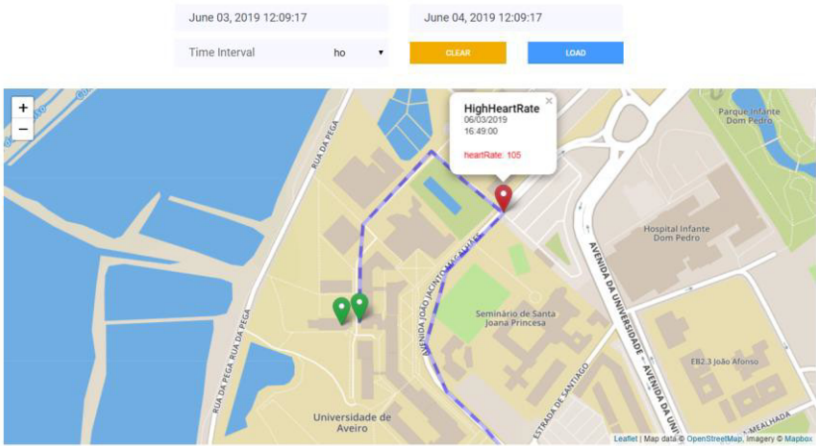


Figure 2. History Page

Users have also available a timeline with relevant events that occur throughout the days. The users can add descriptive markers (annotation) to those events to keep track of their health progress. They include situations like, for instance, heart rate out of the healthy range, presence of high levels of carbon dioxide or other harmful particles, sedentary behaviors (not enough daily steps), few hours of sleep, etc.

Figure 2 presents one of the most emblematic functionalities implemented, the History Page. Here the user can see the places he/she has been, the current location of his/her device(s) and the markers of events that occurred throughout the defined time window. This allows the pairing of events to locations and understand what places

² <https://waqi.info/>

affect him/her the most.

The first version of ContinuousCare application provides 7 main functional areas and presents data related to 20 metrics. Its public API provides 26 public endpoints (i.e. services). The results of the scalability tests executed show that the backend supports comfortably 100 users online, each making 300 requests, ensuring all of the performance requirements established.

5. Conclusions

Modern society is becoming very sensible to the necessity of monitoring environmental conditions and study its impact on personal health. This paper proposes an innovator clinico-environmental system for personal monitoring. ContinuousCare presents itself as a healthcare tool for any person, empowering both people concerned with their own health and medical personnel looking for innovative analysis tools. It has potential to aid any user interested in perfecting their lifestyle and monitoring the surroundings. It combines data collected by air quality monitors with vital signals from personal health devices, and it makes all data collected easy to read, process and consume in an integrated way. Users can also annotate detected events and register daily moods and behaviors. In the future, the efforts will focus on a new system module for data mining aiming to extract key information from events and offer tools to predict diseases and notify users of dangerous behaviors and hazardous environments.

Acknowledgements

This work has received support from the EU/EFPIA Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 806968.

References

- [1] E. O. Tartan and C. Ciflikli (2018). An Android App for Geolocation Based Health Monitoring, Consultancy & Alarm System. IEEE, 978-1-5386-2667-2.
- [2] F. Daniel, F. Casati, P. Silveira, M. Verga and M. Nalin (2011). Beyond Health Tracking, A Personal Health & Lifestyle Platform. IEEE, 1089-7801/11.
- [3] M. Skubic, M. Ma, K. Ai and J. Hubbard (2017). Angel-Echo: A Personalized Health Care Application. IEEE, 978-1-5090-4722-2.
- [4] Y. Chu, Y. Hsieh, C. Wang, Y. Pan, R. Chang (2011). UPHSM, Ubiquitous Personal Health Surveillance and Management System via WSN Agent on Open Source Smartphone. IEEE, 978-1-61284-697-2/11.
- [5] V. Pardeshi, S. Sagar, S. Murmurwar and P. Hage (2017). Health monitoring systems using IoT and Raspberry Pi — a review. IEEE, 978-1-5090-5960-7.
- [6] Colorlib, Hospice Template. <https://colorlib.com/wp/template/hospice/> (accessed in March, 2019).
- [7] Fitbit Co., 'Fitbit Charge 3, Advanced Health and Fitness Tracker'. <https://www.fitbit.com/eu/shop/charge3> (accessed in March, 2019).
- [8] Airboxlab, 'Foobot, the air quality monitor you need'. <https://foobot.io/features/> (accessed March, 2019).

Cohort Creation and Visualization Using Graph Model in the PREDIMED Health Data Warehouse

Christophe CANCÉ^a, Pierre-Ephrem MADIOT^b, Christian LENNE^a,
 Svetlana ARTEMOVA^{bc}, Brigitte COHARD^b, Marjolaine BODIN^{bc},
 Alban CAPOROSSI^{abcd}, Jean-François BLATIER^b, Jérôme FAUCONNIER^b,
 Frédérique OLIVE^b, Daniel PAGONIS^b, Dominique LE MAGNY^b,
 Jean-Luc BOSSON^{abcd}, Katia CHARRIERE^{bc}, Ivan PATUREL^b, Bruno LAVAIRE^b,
 Gabriel SCHUMMER^b, Joseph ETERNO^b, Jean-Noël RAVEY^b, Ivan BRICAULT^{abcd},
 Gilbert FERRETTI^{abf}, Sébastien CHANOINE^{abdf}, Pierrick BEDOUCHE^{abf},
 Emmanuel BARBIER^c, Julien THEVENON^{abf},
 Pascal MOSSUZ^{abf}, Alexandre MOREAU-GAUDRY^{abcd}
^a Univ. Grenoble Alpes (UGA), Grenoble, France
^b CHU Grenoble Alpes (CHUGA), Grenoble, France,
^c Clinical Investigation Centre – Technological Innovation, Inserm, CHUGA,
 UGA, Grenoble, France
^d TIMC-IMAG Laboratory, UGA, CNRS, Vet Agro, Grenoble, France
^e Grenoble Institut des Neurosciences, Grenoble, France
^f IAB, UGA, Inserm, CNRS, France

Abstract. Grenoble Alpes University Hospital (CHUGA) is currently deploying a health data warehouse called PREDIMED [1], a platform designed to integrate and analyze for research, education and institutional management the data of patients treated at CHUGA. PREDIMED contains healthcare data, administrative data and, potentially, data from external databases. PREDIMED is hosted by the CHUGA Information Systems Department and benefits from its strict security rules. CHUGA's institutional project PREDIMED aims to collaborate with similar projects in France and worldwide. In this paper, we present how the data model defined to implement PREDIMED at CHUGA is useful for medical experts to interactively build a cohort of patients and to visualize this cohort.

Keywords. Complex and massive data, medical informatics, health data warehouse, cohorts, data visualization

1. Introduction

Selecting a cohort of patients and gathering associated medical data for research usually involves bringing together information from different sources, potentially isolated from each other in a hospital's information system.

It is thus necessary, for all patients that are likely to be members of a research cohort, to combine their clinical notes with the local information of a clinical department of the hospital, then to enrich them with those distributed in the various hospital databases, to gather all the documents containing demographic, clinical, biological, pharmacological information concerning these patients. This phase often

involves the cooperation with the IT department to transform the potentially complex needs expressed by the research team into querying language to extract the expected information from the hospital's multiple databases. Several iterations of this process between researchers and data managers may be necessary to ultimately lead to the expected data set. However, in this way no interactivity on data exploration is offered to researchers, though it would allow them to obtain or approach faster the best possible result. It is then necessary to organize these data in a coherent and easily explorable form for researchers to select the members of the final cohort that will constitute the data set to be used for the study.

The whole process of collecting the data set associated with a cohort of patients is therefore expensive in terms of mobilizing different health and IT specialists for several months. This paper describes methods and tools implemented as a proof of concept in the framework of the health data warehouse called “PREDIMED” of the Grenoble Alps University Hospital, to enable hospital authorized medical specialists to navigate interactively and in a fluid way through the whole data lake or through their own data set, to create and explore their own cohorts of patients.

PREDIMED is a data lake platform designed to integrate and analyze for research, education and institutional management, the data of patients treated at CHUGA since 1998. PREDIMED contains administrative data (patients and visits), healthcare data (diagnoses, prescription, images, etc.) and, potentially, data from external databases. PREDIMED is hosted by the CHUGA Information Systems Department and benefits from its strict security rules. PREDIMED has several views on the data facilitating its exploration and usage according to the specific needs. In this paper, we specifically present the graph model and the associated tools dedicated to its real-time exploration and data visualization of millions of heterogeneous information.

2. Methods and implementation

Here we describe methods and implementation choices that we used to provide healthcare professionals with an environment allowing them to simply and graphically manipulate the concepts and relationships of their expertise domain to navigate through the hospital's data lake without knowledge of any computer language.

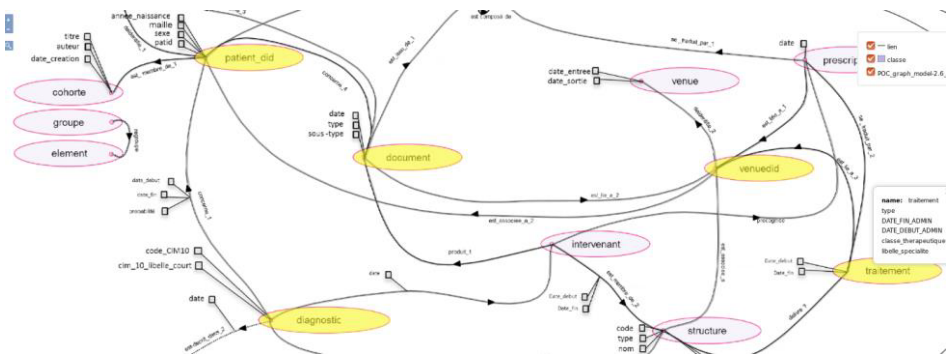


Figure 1: Graph data model of the data lake and interactive cohort construction interface.

The data lake is modeled in an intuitive form that allows each member of the research project, regardless of their profession, to understand its semantics [3]. Figure 1 gives an idea of the graphical visualization of the lake data model. Graph nodes are common medical concepts and graph edges represent the relationships between them.

Patients data from various hospital databases is linked in PREDIMED via an internal key. This technical key is different from the patient number used for care to separate medical information from directly identifying data. The same approach is used for the visit numbers.

Using the graph data model, the team in charge of cohort constitution works with the clinician to build the cohort for the study. The graph model allows starting from any entity since the selection is made by choosing a concept from the model and following an edge that links it to another concept. The traversal can thus start from diagnostic codes (ICD10), periods of hospitalization, administered treatments or any other graph node. The user can then progress in traversing the graph edges by hops, possibly expressing constraints at different levels. Interactively, at each movement through the graph, the system provides the volume of collected data. In Figure 1, one can see that at this intermediate stage, the selected data concerns the entities marked in yellow (here *diagnostics*, *patient_id*, *document*, *venue_id* and *treatment*). The attributes associated with *venue_id* can be filtered (a filter can be set on the hospitalization period (*dt_debut_prestation* for example)).

Behind this graphical selection mechanism is a query language [4] specifically developed for the project that allows the user to query the associated graph database. The query basic hop is built as follows:

“source objects class”[filters]—named relationship[filters]—> “target objects class”, ...

The complete query string that traces the graphical navigation path is associated with the current cohort, which allows the user to relaunch it if the lake content gets richer or to trace what has been done during the exploration phase.

Figure 2 illustrates the expression language of a textual query for the first move through the graph. The presented query is based on a proof-of-concept study building a cohort of patients with diabetes (first move in Figure 2: patients over 20 years old, hospitalized in 2016 and whose biological measurements and received treatments corresponded to the chosen characteristics).

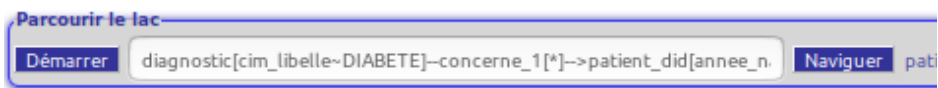


Figure 2: Generated query hop during navigation

3. Results

Through the interactive iterations on this graph, the clinician can visualize the information on the obtained cohort in different ways. Graphical visualization tools associated with the cohort make it possible to highlight the dispersion of patients over the territory or to construct a dashboard with various indicators (see Figure 3).

Based on the same data structure, a visual exploration of the cohort being built provides access to a “360° view” of the cohort elements. The visualization interface in a web browser interacts with the graph database through a software layer (REST API)

that gives an abstraction of the model. This architecture thus limits a strong adherence with the database.

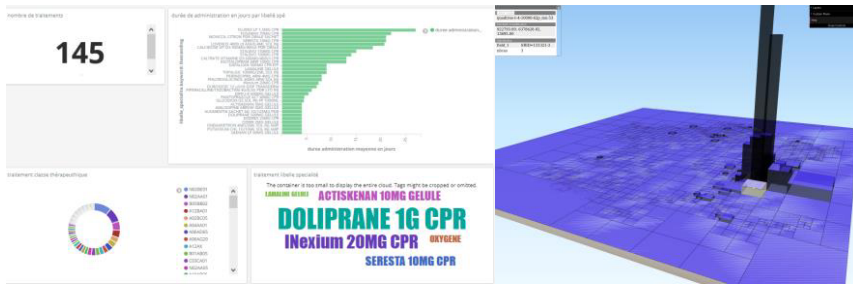


Figure 3: Interactive views describing the current selected data-set

This “360° view” of the patient allows to visualize the different sides of the patient file. The first facet resumes the diagnoses based on the ICD10 codes and the timeline provides the chronology of these diagnoses. Those codes are projected onto a body silhouette highlighting the affected organs or body parts. An oval-shaped representation gives a synthesis of the pathological classes concerned. Other facets provide access to the various documents available in the patient's file. As with diagnoses, a timeline is associated with the list of documents such as drug prescriptions. An equivalent view is available for pharmaceutical prescriptions giving both the list of drugs administered or prescribed (therapeutic class, name of the drug, etc.).

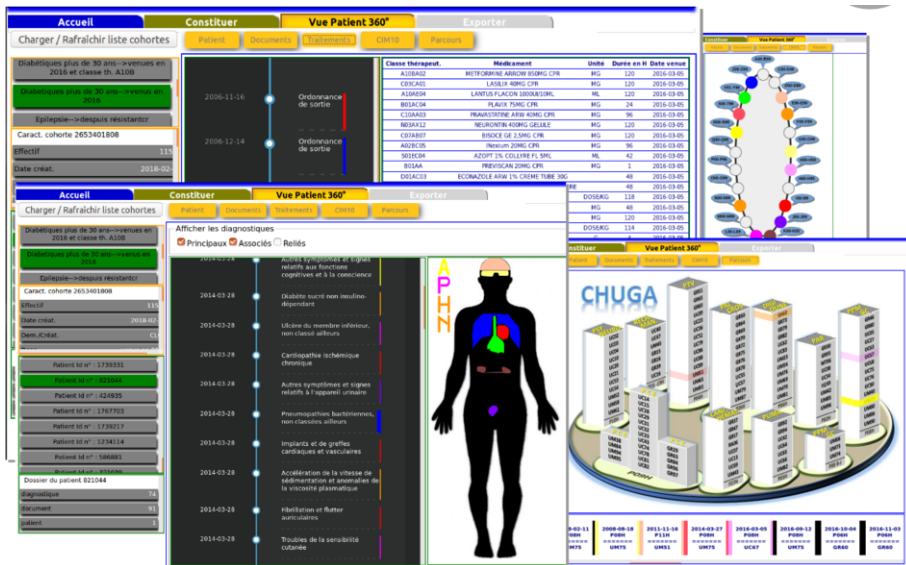


Figure 4: Different views on the patient's file.

4. Discussion

4.1. Flexibility, evolutivity of the model, code genericity.

Graph modeling allows similar concepts to be grouped together and is well-suited to represent flexible and evolving structures. Graph data model implementation with a

graph database is almost straightforward and the high level of manipulated objects is appropriate for the development of generic code, independent of the future data variants.

4.2. Implementation and interoperability

The application environment we have built to operate PREDIMED is based on an open source property graph database, ArangoDB. This in-memory system enables to navigate seamlessly in such complex, highly-connected, massive data organizations. The native JSON-centered data application environment provides both the ability to comply with normative health data standards, to deal with complex objects' attributes (lists, matrices...) and the ability to distribute processing on different and evolutive applications of the PREDIMED architecture. Interoperability is also facilitated by the ability to interconnect with other graphs, using online APIs provided by evolving global data standards for health information such as ICD11.

4.3. Scalability

Unlike a conventional relational database system whose complete request must be written in advance, PREDIMED's alternative request system allows researchers to navigate easily through the data and record the path corresponding to a “built-on-the-fly” request. As the request is executed step by step, even when it is eventually relaunched entirely, there is potentially no limit to the number of steps that constitute the response path to a request. The size of the request and therefore its complexity have no impact on the processing infrastructure.

5. Conclusion

PREDIMED Data Warehouse allows non-IT specialists to use complex and massive data mining tools to effectively and rapidly design patient cohorts for research projects. PREDIMED provides agility and auto-improvement ability in data navigation while making the work traceable, reproducible, and transferable to other data-sets.

This work has been performed as proof-of-concept. The independent supervisory authority in terms of the GDPR (CNIL in France) has authorized the CHUGA to proceed to the operational phase of PREDIMED, on October 10 2019.

References

- [1] Artemova S, Madiot P, Caporossi A, PREDIMED group, Mossuz P, Moreau-Gaudry A, PREDIMED: Clinical Data Warehouse of Grenoble Alpes University Hospital, MEDINFO, Lyon, 2019.
- [2] Birtwell D, Williams H, Pyeritz R, Damrauer S, D. L. Mowery DL, Carnival: A Graph-Based Data Integration and Query Tool to Support Patient Cohort, MEDINFO, Lyon, 2019.
- [3] Bouillé F, The Hypergraph-Based Data Structure: A New Approach to Data Base Modelling and Application, In: Schneider H.J. (eds) GI-7. Jahrestagung. Informatik - Fachberichte, vol 10. Springer, Berlin, Heidelberg, 1977.
- [4] Francis N, Green A, Guagliardo P, Libkin L, Lindaaker T, Marsault V, Plantikow S, Rydberg M, Selmer P, and Taylor A, Cypher: An Evolving Query Language for Property Graphs, in: Proceedings of the 2018 International Conference on Management of Data, ACM, New York, NY, USA, 2018.

Commercial Adoption of AI in the Healthcare Sector: An Exploratory Analysis of S&P500 Companies

Tomasz MUCHA^{a,1}, Olli HALMINEN^a, Henni TENHUNEN^a and Timo SEPPÄLÄ^{a,b}

^a*Department of Industrial Engineering and Management, Aalto University, Espoo, Finland*

^b*Research Institute of the Finnish Economy, ETLA, Helsinki, Finland*

Abstract. The use of Artificial Intelligence (AI) technologies within the healthcare sector is growing. However, there are differences in the speed of commercial adoption of AI across sub-sectors. We employ a dataset including news mentions and executive communications of all S&P500 Health Care Index companies to explore these differences. Pharmaceutical and medicine manufacturing companies had the earliest AI-linked news presence, yet they appeared to be among the slowest commercial implementers of AI. Ambulatory health care services and hospitals, as well as insurance carriers, received media coverage later, but were the quickest to take AI into commercial use. From the theory perspective our results indicate that the classical innovation diffusion theory might not fully explain these differences.

Keywords. Artificial Intelligence, Technology Transfer, Health Technology

1. Introduction

Artificial intelligence (AI) technologies are expected to significantly impact or even transform healthcare provision. Value of goods and services produced by 12 developed economies in the healthcare sector alone is estimated to exceed baseline growth by USD 400 billion in 2035 thanks to the use of AI [1]. However, such value creation and capture can take place only when the technology is translated into action [2]. To understand this transition, information technology (IT) diffusion and assimilation has been one of the central areas of investigation in healthcare informatics [3], information systems [4], and innovation literature [5]. Due to the increasing applications of AI in healthcare, there is a need for studies exploring the status of commercial adoption of AI within that sector.

This paper reports results from an exploratory analysis focusing on the adoption and commercial use of AI technologies by large healthcare companies. We investigate AI and machine learning (ML) related keyword appearances in two contexts for the sample of 62 healthcare sector companies, which were constituents of S&P500 index as of May 2019. The first context is news and media items that relate to the sample companies and include AI or ML keywords. The second context is earnings calls and other investor presentations of the sample companies, where top executives make statements regarding commercial applications of AI and ML technologies.

¹ Corresponding Author: Tomasz Mucha, Aalto University, Finland; E-mail: tomasz.mucha@aalto.fi

2. Data and Methods

Sample of companies selected for this study consisted of S&P500 Health Care Index constituents included in the index as of May 2019. These companies represent some of the largest healthcare companies in the world and are thoroughly covered by news and media, as well as financial analysts. Consequently, there is a large body of publicly available information regarding these companies, which makes a research design leveraging content analysis a feasible approach. The sample companies were divided into six sub-sectors based on their NAICS codes (manufacturing companies were further divided into two subgroups). Following [6], we do not intend to provide explicit definition of AI, but rather rely on implicit definitions embedded in the examples covered in the analyzed content. Nevertheless, our work was guided by an inclusive definition as follows, “Artificial Intelligence . . . is intended to make computers do things, that when done by people, are described as having indicated intelligence” [7].

For each company, we collected data from two sources. News and media items were collected from ProQuest ABI/INFORM Collection using the following search query: *("artificial intelligence" OR "machine learning") AND stype.exact("Magazines" OR "Trade Journals" OR "Newspapers" OR "Reports" OR "Blogs, Podcasts, & Websites" OR "Wire Feeds") AND la.exact("English")*. Transcripts of management presentations were retrieved from Thomson Reuters Eikon database. The types of items searched included not only transcripts of quarterly earnings calls, but also other conferences, media events, financial analyst days and investor events. The case insensitive search query used for that database was: *"artificial intelligence" OR "machine learning"*. The time period covered in this exploratory analysis was 1 January 2004 – 10 May 2019.

Preliminary qualitative analysis focused on the transcripts of management presentations. For each company, the transcripts were analyzed by two researchers to determine whether a commercial application of AI technologies has been mentioned. The definition of commercial application was either (1) reference to a current commercial use of AI or ML technologies (as part of customer offering or internal processes, which are “business as usual”), or (2) commercial launch or implementation utilizing AI or ML technologies in the near future (specific details provided).

The subsequent exploratory analysis had two objectives: (1) to establish the current status of AI use by healthcare sector companies and (2) to inspect the pattern of diffusion of the technology over time with the intention of identifying aspects requiring scrutiny.

3. Results

Nearly all companies (91.9%) included in the sample have been mentioned in news and media in the context of AI, and 43.5% of companies have mentioned commercial use of AI technologies during their investor presentations given by top executives (Table 1).

It appears that news and media coverage in relation to AI and healthcare companies increased sharply starting from 2015. Major increase in number of companies mentioning commercial use of AI during their investor presentations started in 2017. Figure 1 provides more detail on the cumulative percentage of companies covered in news and media items related to AI, as well as those mentioning commercial use of AI.

First mentions of AI in news and media items covering healthcare companies, in many cases, have been far ahead of the recent wave of AI technology implementation by the companies. This is particularly the case for manufacturing companies (sub-sectors 1

and 2). Figure 2 illustrates the publication timings of news and media items. 98% of the items were published during the last four years.

Table 1. Summary statistics for news and media items mentioning sample companies and AI technologies, and top executive investor presentations mentioning AI technologies.

Sub-sector	News and media items			Top executive presentations		
	Count	Mean count per company	Mean date of 1st mention of AI/ML	Count	Mean date of 1st commercial use mention	Share (ratio) of companies with commercial use
1. Manufacturing (excl. Pharmaceutical and Medicine)	721	28.8	2014-11-11	81	2018-08-17	48.0% (12/25)
2. Pharmaceutical and Medicine	1 661	127.8	2013-04-02	14	2019-11-12	15.4% (2/13)
3. Professional, Scientific, and Technical Services	254	31.8	2016-05-04	28	2018-11-27	37.5% (3/8)
4. Ambulatory Health Care Services and Hospitals	327	36.3	2016-12-19	18	2017-10-21	66.7% (6/9)
5. Merchant Wholesalers	148	37.0	2017-09-02	4	2019-11-17	25.0% (1/4)
6. Insurance Carriers	214	71.3	2015-07-16	23	2013-07-09	100.0% (3/3)
Total	3 325	53.6	2015-03-27	168	2018-08-21	43.5% (27/62)

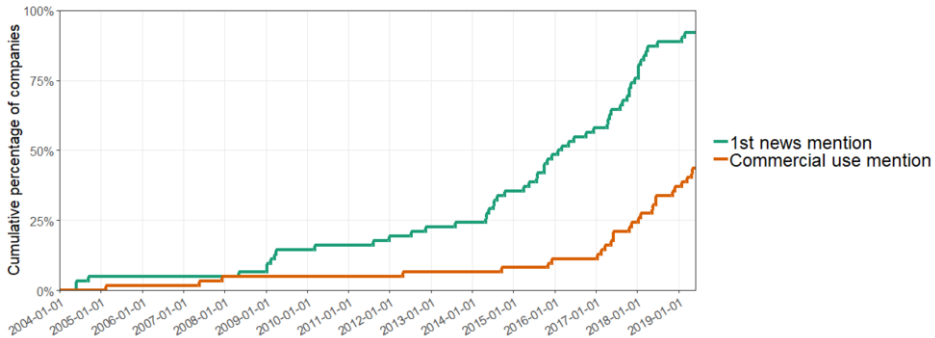


Figure 1. Comparison of cumulative percentage of companies mentioned in AI related news and media items with cumulative percentage of companies mentioning commercial use of AI in executive presentations.

A sub-sector level comparison of cumulative percentage of companies mentioned in news and media in relation to AI, and those whose top executives mention commercial use of AI reveals stark differences. Manufacturing (1 and 2) and Professional, scientific and technical services (3) companies started appearing in news and media items in a relatively even manner over time. The remaining sub-sectors started receiving news and media coverage much later, yet they caught up or even exceeded early-comers. See Figure 3 (left). This pattern is nearly reversed when mentions of commercial uses of AI are considered. See Figure 3 (right). Insurance carriers (6) and Ambulatory health care services and hospitals (4), which received news and media attention later, were not only the first ones to mention commercial uses of AI, but these sub-sectors present also the highest rate of disclosure regarding commercial use of AI. Contrary to the news and

media coverage, Pharmaceutical and medicine manufacturing (2) companies appear to have the lowest rate of commercial utilization of AI.

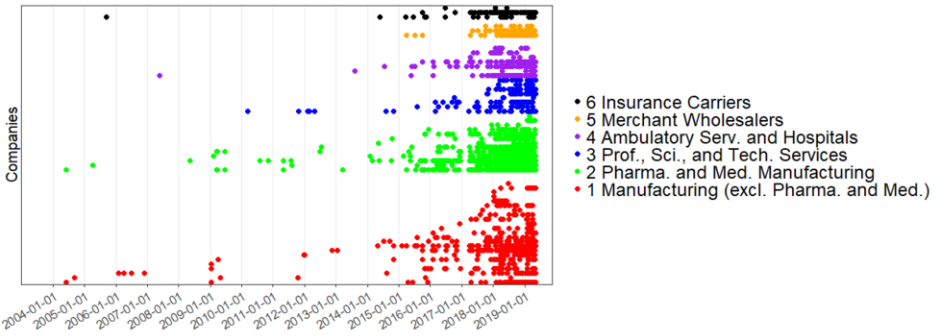


Figure 2. Timing of AI related news and media mentions of healthcare companies. Each row in the chart represents a single company.

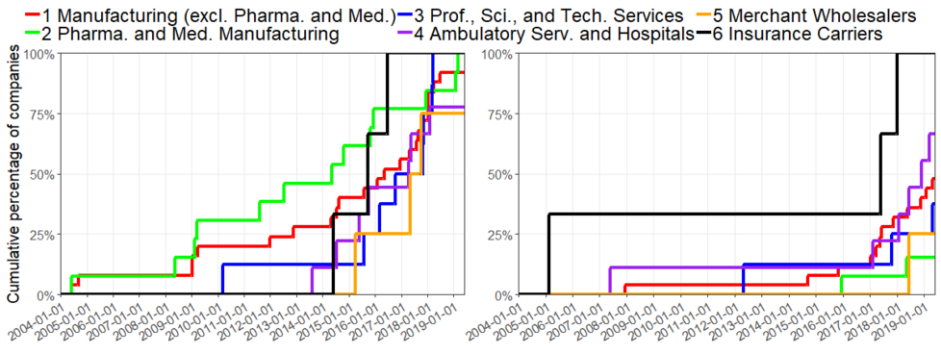


Figure 3. Left: Cumulative percentage of companies mentioned in AI related news by sub-sector. Right: Cumulative percentage of companies with commercial AI mentions in executive presentations by sub-sector.

4. Discussion and Conclusions

Companies from pharmaceutical and medicine manufacturing subsector had the longest and most prominent news and media presence, yet they were among the slowest commercial implementers of AI. Ambulatory health care services and hospitals, as well as insurance carriers received news and media coverage much later, although more rapidly, and were the quickest to take AI technologies into commercial use. Larger media coverage might be a result of pharmaceutical companies having typically a long product developmental timescale, resulting in relatively larger overall fixed costs and possibly also driving larger marketing budgets, thus leading to increased press coverage. AI solutions developed for pharmaceutical companies are also found to attract more venture capital funding than other healthcare-related AI solutions [8].

Classical diffusion of innovations theory (DOI) [5] predicts that new technologies, such as AI, do not spread evenly across the whole population. Instead, a typical cumulative adoption curve takes an “S-shape”— the sample companies taken as a single group appears to follow such pattern (Figure 1). However, sub-sector level analysis and

comparison of news and media coverage (approximating diffusion of information and awareness) with mentions of commercial uses of AI by top executives (approximating diffusion of commercial adoption) indicate that DOI might not fully explain the sub-sector level differences. Disparity between these two diffusion patterns indicates that factors beyond those recognized by DOI might be at play, which is a phenomenon requiring further investigation. From media perspective, the solutions for e.g. insurance carriers might be seen as having less public impact potential than pharmaceutical products [9]. Furthermore, organizational characteristics such as IT intensity and absorptive capacity or use case of AI, such as those illustrated by quotes in Table 2, might explain AI adoption rates [4]. In our future research we intend to further investigate factors driving AI adoption by healthcare companies.

Table 2. Quotes illustrating organizational characteristics and use cases, which might impact AI adoption.

Sub-sector	Quote
1. Manufacturing (excl. Pharmaceutical and Medicine)	“Tempus combines Illumina’s sequencing with machine learning to provide physicians a comprehensive report that highlights key findings to support clinical decision-making.”
2. Pharmaceutical and Medicine	“[...] many different kinds of data sciences we had to bring together. So add to this list graph theory, non-linear machine learning, lot of combinatorial methods, and I’m even going to show you how we’ve used elements of cryptography in some of the analysis...”
3. Professional, Scientific, and Technical Services	“[W]e will leverage [...] machine learning, human genetics and functional genomics to create disease models for NASH and discover relevant drug targets...”
4. Ambulatory Health Care Services and Hospitals	“We use real-time machine learning. We know the demand patterns that are coming into our patient service centers, so we can make real-time staffing decisions and we can take some of the staffing decisions out of the hands of the supervisors and tell them exactly what they’re going to need at that facility next Tuesday at 9:00 in the morning.”
5. Merchant Wholesalers	“[W]e’re basically using machine learning, artificial intelligence to do online claims adjudication, which will make us so much more efficient and so much more scalable...”
6. Insurance Carriers	“[W]e process about 2 billion claims a year, about \$0.5 trillion in medical spend and we transmit about \$130 billion in funds electronically every year and we do so to about 1 million care providers across the country. [...] all of that represents data and [...] increasingly, we’re applying machine learning...”

References

[1] E. Plastino, M. Purdy, Game changing value from artificial intelligence: Eight strategies, *Strategy Leadersh.* 46 (2018), 16–22.

[2] Microsoft CEO: AI can change the trajectory of healthcare if properly used, *Digit. Health.* (2018). <https://www.digitalhealth.net/2018/11/microsoft-ceo-ai-healthcare/> (accessed January 4, 2020).

[3] R.J. Holden, B.T. Karsh, The technology acceptance model: its past and its future in health care, *J. Biomed. Inform.* 43 (2010), 159–172.

[4] R.G. Fichman, The diffusion and assimilation of information technology innovations, *Fram. Domains IT Manag. Proj. Future Past.* 105127 (2000), 105–128.

[5] E.M. Rogers, *Diffusion of innovations*, Simon and Schuster, 2010.

[6] J.K.-U. Brock, F. Von Wangenheim, Demystifying AI: What Digital Transformation Leaders Can Teach You about Realistic Artificial Intelligence, *Calif. Manage. Rev.* 61 (2019), 110–134.

[7] R.A. Brooks, *Intelligence without reason*, Massachusetts Institute of Technology, 1991.

[8] O. Halminen, H. Tenhunen, A. Heliste, T. Seppälä, Factors Affecting Venture Funding of Healthcare AI Companies, *Stud. Health Technol. Inform.* 262 (2019), 268–271.

[9] K. Viswanath, K.D. Blake, H.I. Meissner, N.G. Saiontz, C. Mull, C.S. Freeman, B. Hesse, R.T. Croyle, Occupational practices and the making of health news: A national survey of US health and medical science journalists, *J. Health Commun.* 13 (2008), 759–777.

A Computational Adverse Event Detection Matrix

Mary GADDE ^a and Melody PENNING ^b

^a University of Arkansas, Little Rock, Arkansas, United States ^b Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, United States

Abstract. Harms caused during healthcare encounters are pervasive and occur at an alarming rate; therefore, building a set of computational detection methodologies in the adverse event area is urgently needed to address this problem. To understand the entire range of adverse event detection methods currently in practice we have developed a computational adverse event detection matrix. This structure is made of methods used presently at US hospitals to detect patient safety events. It contains adverse event 1) concepts and 2) synthesized detection strategies as well as calculations of overlap of coded data in the subset of algorithms implemented completely computationally. Most importantly, this matrix provides a clear picture of coverage gaps in the detection of adverse events.

Keywords. Patient safety, Data quality

1. Introduction

Harms or Adverse events (AE)s caused during healthcare encounters are pervasive, have a significant impact on patient wellbeing, and occur at a staggering rate. Specifically, one of every ten hospitalized patients develops a healthcare-acquired condition, and twelve million outpatients experience a diagnostic error, which may result in harm, every year. [3] AEs can be categorized as unavoidable (e.g. part of the patient's condition), potential (near miss or never events) or preventable and can come from active medical errors at the point of care or latent causes present within care settings. [9; 13] The Institute for Healthcare Improvement uses a definition for AEs as "Unintended physical injury resulting from or contributed to by medical care (including the absence of indicated medical treatment), that requires additional monitoring, treatment, or hospitalization, or that results in death".

This study is designed to determine which of the events designated as harms, or AEs, are currently 1) identifiable algorithmically, 2) which are identifiable using other methods, and 3) which are likely going undetected due to the methodology used. AEs are typically tracked and quantified using incident reporting, medical record review or claims data, although other methods are also available, as shown in Table 1. None of these processes captures all AEs.[10] Comparative studies suggest that the most effective methodology is the Medical record review (MRR).[9] As this process is time-consuming, efforts have been made to increase the efficiency of this work with instruments like the Institute for Healthcare Improvement (IHI) Global Trigger Tool. [3; 7]. To ensure as many AEs are detected as possible while at the same time not overloading healthcare staff automation has been introduced in several of the methodologies listed in Table 1. Detecting AEs using rule-based algorithms makes automation possible. [8] But this

automation is only as good as the combination of the phenotype definition and the data it consumes. Not all AE will be detectable using rule-based algorithms. Active errors occur at the point of contact between the patient and the healthcare system and are typically observable whereas latent errors are process errors, where no single event is the source of the problem. AEs arising from active errors are more detectable than those arising from latent causes, and direct process observation or data mining are the best solutions to finding these types of situations. In general, the methods listed in Table 1 can miss as much as 90% of all AE types. [5]

Table 1. AE Detection Methods Currently in Use

Medical Record Review (w/ and w/out Trigger Tools)
Studies based on interviews w/ healthcare providers
Direct Observation
Incident Reporting Systems
External Audit
Studies of legal claims and complaints
Administrative Data
Computational Methods using EHR
Autopsy

We are currently in need of common terminology to quantify the occurrence of and detection methods for AEs. In the US, governmental organizations and Patient safety organizations (PSO), regulated by the US Agency for Healthcare Research and Quality (AHRQ), collect data and do analysis and reporting on patient safety concerns. Many of these organizations have designed approaches to describe, classify and detect patient safety events. Methods used include retrospective surveillance with administrative coded data, observation and voluntary reporting as well as MRR. All of these methods are flawed in that they miss AEs or are excessively expensive so cannot be conducted at scale. [12; 14] Additionally, large scale detection methods are not sensitive enough to show improvement over time. [3]

In an effort to find common ground among the methods currently in practice, we have developed a computational AE detection matrix. This matrix provides a clear picture of what is available and what is missing in the detection of AEs. As part of this process we have also investigated the overlap in computationally implemented algorithms.

2. Methods

In order to understand where there is overlap in current AE detection methodologies and to discover which AE are not the subject of *in practice* methodologies, we have classified currently accepted PSO AE detection algorithms into the International Classification for Patient Safety categories from the World Health Organization (WHO). Patient safety information is categorized into standardized concepts that represent descriptions that are valid internationally and can be used for comparison, measurement, monitoring, analysis and interpretation of information for improved patient safety.[15] This framework breaks patient safety events into ten high-level classes which are broken down into very specific but generalizable concepts two of which are directly comparable to the PSO terminology, incident type and patient outcomes. Although, some incidents may be classified in more

than one category in the WHO framework we have kept to a one to one relationship with detection algorithm.

Right now, there are a few standard approaches to AE detection and none are both highly efficacious and also computational. Some are completely computational and can detect roughly half of AE [1], and some use MRR, which is very thorough but limits the number of cases evaluated due the time required. In previous studies commonalities among detection strategies were synthesized and found to be predominantly focused in the areas of surgery and care management.[11] This study seeks to extend the previous work by including areas in healthcare where AE are recognized, but no external detection method is currently in place to identify them. This work includes the extraction of the rule-based definitions. For the computational approaches, the extraction of International Statistical Classification of Diseases, Injuries and Causes of Death (ICD) codes used in the detection of AEs from organizational source documents was necessary to build an understanding of the degree of overlap between computational methodologies. Commonalities have arisen between the ICD codes coming from government bodies The Centers for Medicare and Medicaid Services (CMS) and PSOs (AHRQ). AEs are detected for Patient Safety Indicators (PSI) coming from AHRQ and Human Acquired Conditions (HAC) coming from CMS.

The manual and hybrid approaches to AE detection were extracted from the organization website sources. These cannot be compared directly with the computational methods in terms of data points due to the incorporation of human judgment. The most effective method with respect to efficiency as well as thoroughness is the hybrid approach. Together, review of medical records with computerized monitoring capture conditions that are associated with more preventable and potential AEs than any singular method. [4; 6; 9]

The sources used for these algorithms are (a) CMS, hospital-acquired conditions, (b) IHI, Global Trigger Tool (c) The Agency for Healthcare Research and Quality (AHRQ, PSI), (d) The Joint Commission (TJC, Sentinel Event definitions) and (e) The National Quality Forum (NQF, serious reportable event (SRE) definitions).

3. Results

Both the CMS HAC codes and the AHRQ PSI rates use ICD codes to detect patient harm. As shown in Figure 1 - 5,702,505 ICD codes are used, and 70,384 are shared. AHRQ uses a large portion of the CMS set plus a great deal more.

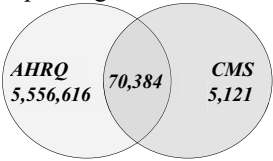


Figure 1. AHRQ and CMS detections methods which use ICD codes exclusively.

This large overlap exists due to the use of the Medicare Severity Diagnosis Related Group (MS-DRG) system for financial considerations by the Centers for Medicare and Medicaid Services. This system, in effect, determines the definition of an AE by associating payment penalties with the code combinations that represent harm conditions. The idea of patient safety and the prevention of AEs was the impetus for the MS-DRG

system. That being the case, any computational AE detection methodology in the US will need to also incorporate information from this system.

The AE detection matrix shown in Table 2 classifies rule-based algorithms according to the WHO Patient Safety Conceptual Framework. The WHO incident types are in the leftmost column. The binary patient outcome severity, listed as present or not, is across the top. Both CMS and AHRQ use the MS-DRG system which provides grouping information about severity of an inpatient stay in order to facilitate payment for services. The methodology used in detection is the column title in Table 2 listing the spectrum of methodologies in

Table 2. Computational AE Detection Matrix

Level of Harm Specified		No	Yes		
		EHR Data	EHR Data	Trigger w/ MRR	Observation or MRR
WHO Incident Categories	Behavior				NQF/TJC
	Blood/Blood Products		AHRQ/CMS	IHI	NQF/TJC
	Clinical Administration	AHRQ		IHI	NQF/TJC
	Clinical Process/Procedure	AHRQ	AHRQ/CMS	IHI	NQF/TJC
	Healthcare Associated		AHRQ/CMS	IHI	NQF
	Infection				
	Infrastructure/Building/Fixtures				NQF
	Medical Device/Equipment				NQF
	Medication/IV Fluids			IHI	NQF
	Oxygen/Gas/Vapor			IHI	NQF
	Patient Accidents		AHRQ/CMS	IHI	NQF/TJC

Increasingly difficult and time consuming to complete → practice from completely computational to completely manual.

4. Discussion

This work lays out detection strategies currently in use against an international conceptual framework of AEs, thereby exposing the sparseness of computational solutions in today’s healthcare environment. Each of the fundamental patient safety events in the framework has many subtypes, and out of these 10 categories only 5 are computationally implemented, and one of these doesn’t indicate severity. That leaves half of AE types to voluntary reporting and record review with no external means of being discovered.

Moreover, the computationally implemented detection algorithms rely on coded administrative data. There are concerns about accuracy in this type of data because not all cases of a complication will be captured, leading to both false positives and negatives. Lastly, data quality is a problem in any data source and missing information is the most prominent issue in healthcare data resulting in codes that may not fully reflect the clinical case.[2]

While gaps exist in computational methods, there is good coverage in the application of medical record review. This approach addresses the false positive and false negative concerns but it is intensely laborious and not practical for truly safeguarding patients. The combination of review using a data-driven trigger tool approach has potential as well but is not yet in common practice due to the complexity of the process. [12]

5. Conclusion

Harms caused during healthcare encounters are ubiquitous and occur at an alarming rate and building a complete set of computational detection methodologies in the AE space is urgently needed to address the problem. This is a many-dimensional problem due to multiple organizations, terminologies, detection methodologies and levels of harm. The matrix of patient safety events developed in this study and the resulting synthesis of associated detection algorithms provides a clear picture of areas where coverage gaps exist in AE detection.

References

- [1] S. Ackroyd-Stolarz, et al., Validating administrative data for the detection of adverse events in older hospitalized patients, *Drug, healthcare and patient safety* **6** (2014), 101.
- [2] E.S. Berner, et al., Data quality in the outpatient setting: impact on clinical decision support systems, *AMIA Annu Symp Proc* (2005), 41-45.
- [3] D. Berwick, et al., Free from harm: Accelerating patient safety improvement fifteen years after To Err Is Human, *Boston, MA National Patient Safety Foundation* (2015).
- [4] V. Bhise, et al., An electronic trigger based on care escalation to identify preventable adverse events in hospitalised patients, *BMJ Qual Saf* **27** (2018), 241-246.
- [5] D. C Classen, et al., 'Global Trigger Tool' Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured, 2011.
- [6] J. Davis, et al., The Accuracy of Trigger Tools to Detect Preventable Adverse Events in Primary Care: A Systematic Review, *J Am Board Fam Med* **31** (2018), 113-125.
- [7] R. Frances A. Griffin, MPA: Director, IHI Roger K. Resar, MD: Senior Fellow, IHI, IHI Global Trigger Tool for Measuring Adverse Events, in: *Innovation Series 2009*, 2009.
- [8] Q. Li, et al., Phenotyping for patient safety: algorithm development for electronic health record based automated adverse event and medical error detection in neonatal intensive care, *Journal of the American Medical Informatics Association : JAMIA* **21** (2014), 776-784.
- [9] P. Michel, Strengths and weaknesses of available methods for assessing the nature and scale of harm caused by the health system: literature review, *World Health Organization* (2003).
- [10] I. Mitchell, et al., Patient safety incident reporting: a qualitative study of thoughts and perceptions of experts 15 years after 'To Err is Human', *BMJ Quality & Safety* **25** (2016), 92-99.
- [11] M. Penning, et al., A Qualitative Evidence Synthesis of Adverse Event Detection Methodologies, *Studies in health technology and informatics* **257** (2019), 346-351.
- [12] C. Sammer, et al., Developing and Evaluating an Automated All-Cause Harm Trigger System, *The Joint Commission Journal on Quality and Patient Safety* **43** (2017), 155-165.
- [13] D.F. Sittig,H. Singh, Defining health information technology-related errors: new developments since to err is human, *Arch Intern Med* **171** (2011), 1281-1284.
- [14] T.M. Waters, et al., Use of ICD-9-CM Codes to Identify Inpatient Fall-Related Injuries, *Journal of the American Geriatrics Society* **61** (2013).
- [15] G. World Alliance For Patient Safety Drafting, et al., Towards an International Classification for Patient Safety: the conceptual framework, *Int J Qual Health Care* **21** (2009), 2-8.

Computer-Assisted Quality Assessment of Aortic CT Angiographies for Patient-Individual Dose Adjustment

Marja FLEITMANN^{a,1}, Kira SOIKA^a, Andreas Martin STROTH^b, Jan GERLACH^b,
Alexander FÜRSCHKE^b, Peter HUNOLD^b, Jörg BARKHAUSEN^b,
Arpad BISCHOF^{b,c} and Heinz HANDELS^a

^a *Institute of Medical Informatics, University of Lübeck*

^b *Department of Radiology and Nuclear Medicine, UKSH Lübeck*

^c *IMAGE Information Systems Europe, Rostock*

Abstract. Iodine-containing contrast agents (CA) are important for enhanced image contrast in CT imaging especially in CT angiography (CTA). CA however poses a risk to the patient since it can e.g. harm the kidneys. In clinical routine often a standard dose is applied that does not take differences between individual patients into account. We propose a method that as a preliminary stage determines excessive image contrast and CA overdosing by assessing the image contrast in CTA images obtained with the ulrich medical CT motion contrast media injector with RIS/PACS interface. A resulting CA dose recommendation is linked to a set of clinical parameters collected for each assessed patient. We used the established data set to implement an automatic classification for individual CA dose adjustment. The classification determines similar cases of new patients to take on the associated CA dose adjustment recommendation. The computation of similar patient data is based on the previously collected patient-individual parameters. The study shows that as basis for a recommendations the largest proportion of patients receive too much CA. A first evaluation of the automatic classification showed an overall error rate of 22% to recognize the correct class for CA dose adjustments using a k-NN-Classifer and a leave-one-out method. The classification's positive predictive value for correctly assigning a CA overdosing was 85.71%.

Keywords. CT Angiography, Contrast Agent, Personal Health Care

1. Introduction

Radiological examinations using contrast agents (CA) belong to the basic methods of image diagnostics and interventional therapy. CA used in CT angiography (CTA) gains the contrast enhancing ability through the contained iodine. Iodine-containing CA is however known for being harmful to the renal system with contrast-induced nephropathy being the third leading cause of hospital acquired acute renal failure [1]. However, a standard CA dose is administered to patients in clinical practice. A few methods were proposed to simulate the CA flow based on hemodynamic models [2, 3] that consider

¹ Marja Fleitmann, Institute of Medical Informatics, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; E-mail: fleitmann@imi.uni-luebeck.de.

physiological parameters of individual patients. Nevertheless, the focus there is on the optimization of the CTA scan and not on the determination of a patient-individual CA dose. In [4] a method was proposed to adjust the CA dose based on the patient's body weight and the CT's tube potential. In our approach we propose a method to perform an automatic quality assessment of CTA data (Fig. 1 (blue box)) as a ground truth for CA dose adjustment recommendations for individual patients. Similar to [5, 6] case-based reasoning a Nearest-Neighbor approach is applied, since it showed convincing results in tasks with medical backgrounds and is transparent to a high degree. The automatic classification based on the k-NN uses a multitude of patient-individual parameters to determine the most similar already assessed patient data for a particular case and takes over the corresponding CA dose adjustment recommendation (Fig. 1 (yellow box)).

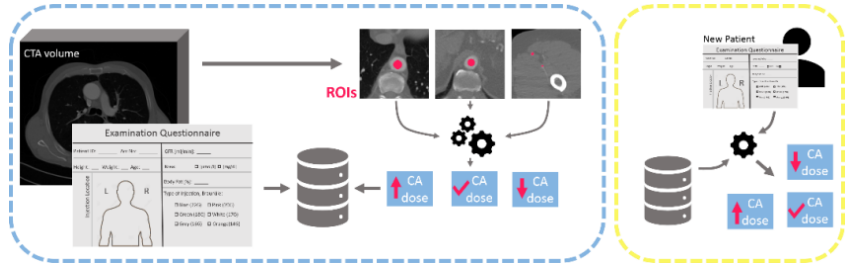


Figure 1. Quality assessment (blue box) and automatic classification (yellow box).

2. Study

The required CTA data is sourced from a study at the radiology department of the UKSH Lübeck. The goal of the acquiring stage is to collect 400 CTA data sets evenly subdivided into four scan protocols: aortic angiography, pulmonary angiography, pelvic-leg angiography and abdomen angiography. In addition, a multitude of patient-individual parameters are collected. These include parameters like age and height, laboratory parameters (e.g. creatine content) and non-invasive physiological parameters (e.g. blood pressure). To this state 226 CTAs featuring 75 aortic angiographies as the largest share were transferred. Therefore, in this paper we focused on the analysis of CTA data with the aortic scan protocol.

3. Methods

3.1 Quality Assessment of Image Contrast

To enable CA dose adjustments an automatic quality assessment of existing CTA images is implemented. The image contrast is assessed based on selected regions of interest (ROI) in diagnostically relevant locations affected by the CA. The assessment serves as a classification implying whether the applied CA dose is optimal or can be decreased. Three ROIs for each aortic CTA image are evaluated and were defined with the aid of radiological experts:

- ROI 1: Aorta, level: pulmonary artery bifurcation
- ROI 2: Aorta, level: front or lateral outflows of the aorta respectively

- ROI 3: Vessels of the periphery, level: last slice of CTA image

At first the medical practitioner must choose a slice of interest (SOI) for each ROI. Then, an algorithm is implemented to automatically detect the aorta and the peripheral vessels in the SOIs, respectively. To locate the aorta for ROI 1 and 2 an algorithm based on the Circle Hough Transformation (CHT) [7] was used. Based on the resulting Hough space we made the empirical decision to consider the six most probable circles. Because there are more circular shaped structures e.g. the vertebral body, a cost function was implemented similar to [8] to identify the aorta. The cost function assesses the variance, differences in radii and distances of the centers of multiple consecutive circles positioned in adjacent axial slices with a dynamic programming approach. The radius of the starting circle that accumulates the lowest costs is decreased by four pixels to evade calcifications and is declared as the ROI. For detection of ROI 3 the vesselness filter by [9] is applied to the last four slices of the CTA Image to mark the peripheral vessels. The vesselness filter can extract bright lines in 3D based on the Hessian Matrix and their eigenvalues. To remove partially segmented bone structures and calcifications a thresholding with the values [660, 3100] Hounsfield units (HU) was applied to the last slice of the CTA image and subtracted from the corresponding filtered slice to determine the ROI 3. Each ROI is then assigned a category from Tab. 1 with respect to the mean HU value of the particular ROI. The thresholds for the target area were similarly chosen to [10].

Table 1. Definition of the categories of image contrast. The range refers to the mean HU value in the ROI.

Category	Range [HU]	Definition
A	≤ 180	HU value too low
B	181 – 240	Lower tolerance area
C	241 – 300	Target area
D	301 – 360	Upper tolerance
E	>360	area Excessive HU value

Aided by radiological experts rules were formulated to describe the image quality as ground truth for the automatic classification. Applying the rules to the combination of the three ROIs results in one of the three possible classes from Tab. 2.

Table 2. The quality assessment results in one of the following classes and recommendations.

Class	Definition	Rule
1	Image contrast insufficient	At least one ROI is in category A
2	Optimal CA dose	Left over combinations
3	Decrease CA dose	At least one ROI is in category E, the two remaining are at least in category D

3.2 Automatic Classification

The main objective of our project is to enable patient-individual adjustments of the CA dose based on the previously classified CTA data. As a mean to achieve this a k-Nearest Neighbor classification was implemented as a first approach which is suitable if only small training sets are available. The k-NN classification identifies the *k* most similar already classified patient data. The according class is then adopted using majority vote and the recommendation for the dose adjustment can be applied. As features nine clinical parameters were considered based on their previously assessment in literature [2]: age, weight, height, body mass index (BMI), gender, glomerular filtration rate (GFR), ankle brachial index (ABI), oxygen saturation and blood pressure (systolic). For gender we assumed a non-ordinal distribution however further possibilities will be considered.

4. Experiments and Results

4.1 Data

For our experiment we used 75 CTA volumes scanned with an aortic angiography protocol. The CA was administered with a programmed injection rate of 5 ml/s.

4.2 Quality Assessment – Experiments and Evaluation

To test the algorithm's ability to locate the aorta in the selected slice the number of cases where the aorta was manually corrected was considered. The same approach was used to evaluate the vesselness filter to determine ROI 3.

4.3 Quality Assessment – Results

69 cases were processed successfully. The aorta was located correctly in 71.01 % cases. The vesselness filter was corrected in 50.72 % cases. Using the definition of the rules six CTA volumes were assigned to class one, 18 to class 2 and 45 to class 3. As a first insight the quality assessment confirms that there is a CA overdosing in clinical routine examinations. The algorithm proved to be successful for detecting the aorta rarely resulting in failures mostly when confronted with major aortic deformations and dissections (s. Fig. 2). The correction rate of the vesselness filter for ROI 3 was rather high but given the variance in the CTA data, the smallness of the peripheral vessels and the high effort caused by manually determining the vessels, the use of the filter was still rated as a success.

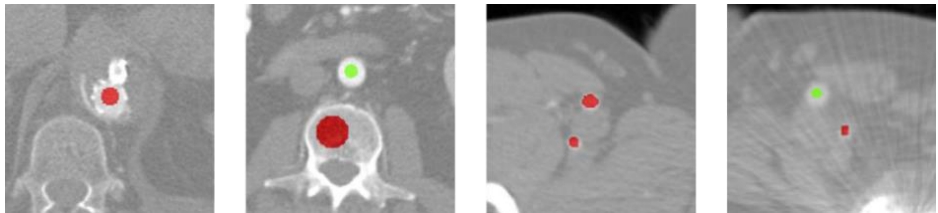


Figure 2. Images of the found aorta and peripheral vessels (red) and the corrections (green) respectively.

4.4 Automatic Classification – Experiments and Evaluation

For the k-NN classification all 2^9-1 combinations of normalized clinical parameters were tested with $k \in [1, 3, 5, 7]$. We used leave-one-out validation to evaluate each classifier resulting in a failure rate. In addition, the positive predictive value for classifying a CA overdose was determined. The Euclidean and the Mahalanobis metric were used as distance measures.

4.5 Automatic Classification – Results

The lowest overall failure rate was 22 % meaning that in 47 cases the determined class matched the prior classification. The rate was achieved with the Mahalanobis metric as

the distance measure and $k = 3$ neighbors. The optimal combination of clinical parameters used was: ABI, height, BMI, oxygen saturation and blood pressure.

Considering the clinical parameters some (BMI, height) were expected due to their examined correlations with contrast enhancement in literature [2, 4]. Others like ABI, oxygen saturation and blood pressure can be considered as indicators for the vessel condition and therefore influence the CA flow and distort the CTA scan. The positive predictive value of the automatic classification after merging class 1 and 2 to correctly assign class 3 (Tab.2), which is crucial for dose reductions in the use case is 85.71 % and showed a rather high prediction quality.

5. Discussion and Conclusion

In this paper we presented a method for an automatic quality assessment of CA-induced contrast enhancement in CTA images with an application for patient-individual dose adjustment. By considering clinical parameters of a new patient the examiners could decrease the CA dose in the use case to lower the risks of CA-induced side effects. While the implemented algorithm is robust for locating the aorta the need of user-determination for the SOI could also be automated. Furthermore, it is debatable how feasible it is to choose the last available slice for the ROI. Therefore, a fixed landmark for the ROI 3 e.g. the aorta femoralis communis will be set. As a next step we will consider deep learning approaches to replace the k-NN classifier for more precise results based on the increased number of patients in our ongoing study. In addition, the presented system will be applied to returning patients for validation.

References

- [1] Pannu, N., Wiebe, N., Tonelli, M., and Alberta Kidney Disease Network. (2006). Prophylaxis strategies for contrast-induced nephropathy. *Jama*, 295(23), 2765-2779.
- [2] Bae, K. T. (2010). Intravenous contrast medium administration and scan timing at CT: considerations and approaches. *Radiology*, 256(1), 32-61.
- [3] Cheng, Y., and He, W. (2016). Establishment of the intracranial hemodynamic model based on contrast medium and clinical applications. *Medicine*, 95(49).
- [4] Vasconcelos, R., Vrtiska, T. J., Foley, T. A., Macedo, T. A., Cardona, J. C. M., Williamson, E. E., McCollough, C. H., and Fletcher, J. G. (2017). Reducing iodine contrast volume in CT angiography of the abdominal aorta using integrated tube potential selection and weight-based method without compromising image quality. *American Journal of Roentgenology*, 208(3), 552-563.
- [5] Chuang, C. L. (2011). Case-based reasoning support for liver disease diagnosis. *Artificial Intelligence in Medicine*, 53(1), 15-23.
- [6] Bichindaritz, I., and Marling, C. (2006). Case-based reasoning in the health sciences: What's next? *Artificial intelligence in medicine*, 36(2), 127-135.
- [7] Kimme, C., Ballard, D., and Sklansky, J. (1975). Finding circles by an array of accumulators. *Communications of the ACM*, 18(2), 120-122.
- [8] Kurkure, U., Avila-Montes, O. C., and Kakadiaris, I. A. (2008, May). Automated segmentation of thoracic aorta in non-contrast CT images. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 29-32). IEEE.
- [9] Sato, Y., Nakajima, S., Atsumi, H., Koller, T., Gerig, G., Yoshida, S., and Kikinis, R. (1997, March). 3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In *CVRMED-MRCAS'97* (pp. 213-222). Springer, Berlin, Heidelberg.
- [10] Becker, C. R., Hong, C., Knez, A., Leber, A., Bruening, R., Schoepf, U. J., and Reiser, M. F. (2003). Optimal contrast application for cardiac 4-detector-row computed tomography. *Investigative radiology*, 38(11), 690-694.

Continuous Monitoring and Statistical Modelling of Heart Rate Variability

Adina NITULESCU^{a,1}, Mihaela CRISAN-VIDA^a and Lacramioara STOICU-TIVADAR^a

^a*Faculty of Automation, and Computers, University Politehnica of Timisoara, Romania*

Abstract. The paper presents a sensor network system implemented to continuously monitor and detect cardiovascular diseases experienced in patients at high-risk. The device is designed to capture and transmit the ECG signals and pulse to the patient's mobile phone. Later interpretation of the single-lead ECG signals will result into a HRV analysis and heart-rate detection that is compared with the pulse oximeter results for validation. The fastest alert will be issued to doctors, relatives and hospitals, using the proposed data processing algorithm implemented in the patient's mobile phone. The continuous usage of this monitoring system provides an effective, efficient and fast health care service to patients at risk, even if the doctors or relatives are not near the patients.

Keywords. sensor, electrocardiogram, alert, emergency, wavelet transform

1. Introduction

In the past, it was difficult to use sensors and front-end electronics in wearable technology to gather physiological and movement data because of their size. But now, wearable sensors are available at a much lower cost and utilized in digital health monitoring systems with miniature circuits.

Prevention is especially critical for cardiovascular diseases and electrocardiogram (ECG) is the most undisputed and widely accepted tool to detect and diagnose them. Apart from their enormous impact in older people life expectancy, cardiovascular diseases are also the main cause of death for the population among 44 and 64 years and detecting their symptoms in time is critical to avoid irreparable damages or death. Our solution comprises of an integrated system module that captures the physiological ECG and SpO2 data that is measured using a single board computer card, a micro-controller and an e-health sensor platform.

Timely access to emergency cardiac care and survival is partly dependent on early recognition of heart-attack symptoms and immediate action by calling emergency services. In a 2016 study [1], most persons (95%) recognized chest pain as a heart-attack symptom, but only 11% correctly classified all symptoms and knew to call 1-1-2 when someone was having a heart-attack. This proved the need of medical devices that continuously monitor and record the physiological parameters of patients outside the clinical environment as part of a care pathway.

[2] presents a mobile healthcare application for self-diabetes management. The application synchronizes data with hospital's EMR database to provide accurate data.

[3] presents a mobile app integrated into a hospital workflow to achieve medication list portability via Fast Healthcare Interoperability Resources.

[4] presents FHIRbox which is a distributed system under development by the authors of the paper for integrating data from various diagnosis devices and uses FHIR for exchanging clinical information.

Still an important issue is the continuity of healthcare, the communication standards will lead to a better patient healthcare activity. The exchange of information between different medical unit's information systems and devices, and understanding each other means that the interoperability request is satisfied. In this paper the standard which will be used is HL7 Fast Healthcare Interoperability Resources (FHIR) [5]. This standard provides access to medical data via a RESTful API which uses XML and JSON. The benefits of using FHIR are: real-time, query-able, patient – centric interface to EHR Resources; HTTP/REST greatly strengthens interoperability, the specification of this standard provides all the needed data and modern code libraries and architecture.

2. System description

The system is designed to capture the physiological ECG signals and detect the heart-rate using wavelet decomposition and fixed thresholding of wavelet coefficients. The following components were integrated in the system capturing module:

- Pulse Oximeter module, designed to capture patient-specific pulse and saturation oxygen signals, digitize and send them to the e-Health Sensor shield.
- ECG electrodes, diagnostic tools that are routinely used to assess the electrical and muscular functions of the heart further useful in the diagnosis of a myriad of cardiac pathologies.
- GSM shield: In order to make calls on threshold events, the system includes a GSM shield for the Arduino Uno, connected directly throughout the pins.
- Arduino + Raspberry Pi module, connected by using the I2C communication protocol.

The captured data is further processed for diagnosis, treatment or both. It also enables the patient to be more engaged and auto-monitor their health status by regularly or continuously checking the values of the vital parameter stated above. The data collected is stored in a database which further can be accessed by an API for creating a file in the HL7 FHIR standard format. The API used in this system is Azure API for FHIR [6]. The data sent using FHIR for our particular case is accessed by the Obstetrics-Gynecology Department Information System previously developed and in use by the medical staff [7]. The Obstetrics-Gynecology Department Information System (OGD IS) was developed using Visual Studio.NET 2015, ASP.NET pages and C# language, it is a responsive web application, and may be accessed by the doctors using a tablet. OGD IS has a statistical component which allows the medical staff tracking of health parameters and draw different charts related to treatments, or IoT systems.

Figure 1 presents the general architecture of the solution presented in this paper.

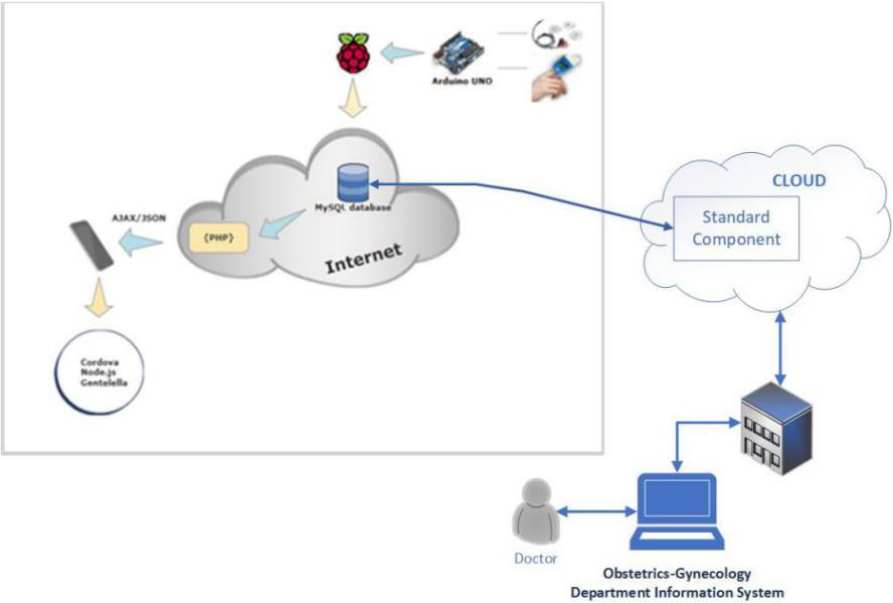


Figure 1. Overview of the system

2.1. Mobile application

The client mobile application is designed specifically to be accessed from personal smart-phones and it permanently displays the patient status and time evolution illustrated by graphs and alarms in case of threshold events. The application is built using PHP, HTML, Ajax and JavaScript technologies. This client application can be easily accessed from mobile Android devices to provide healthcare professionals with the ability to check their patients' status when not near them, and can receive alarms when a patient's condition is critical.

In figure 2 we present the mobile application user interface, called RpiHm (Raspberry Heart Monitor) build using Apache Cordova, an open-source mobile development framework. The template used is Gentelella, a free to use Bootstrap admin template. It uses Bootstrap 3 styles along with a variety of powerful jQuery plugins and tools to develop a powerful framework for creating admin panels and back-end dashboards.

The app can be downloaded on the phone and used by both the patient and the physician for continuous monitoring. It displays information regarding the beats-per-minute, the oxygen saturation (SpO2) and a graph. Upon launching, the ECG graph begins collecting data from the MySQL database and constructs the waveform.

3. Statistical analysis

By loading the data set into Tableau, we can perform a statistical analysis and derive models and patterns in our values. In Figure 3 we can see the preprocessed ECG data that is being used to detect the heart-beat rate. The detection is performed by using discrete wavelet transform (DWT) in order to obtain the amplitude and duration of the

characteristic ECG waves and to determine the RR interval time series. The programming language used for the heart-rate variability (HRV) analysis is R, a free software environment for statistical computing and graphics.

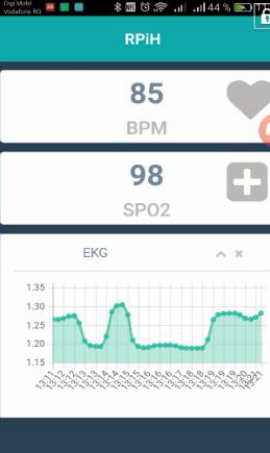


Figure 2. Mobile app

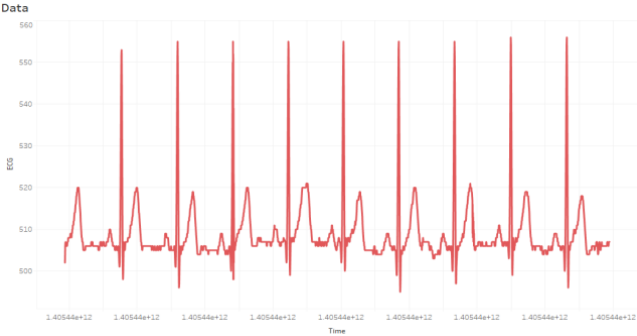


Figure 3. Preprocessed ECG data

By further refining, we conclude that roughly all the values from our dataset are situated in the 70-80 bpm interval, a healthy distribution since any resting value outside this range can be considered a threat. We further want to predict the probability of having a heart disease and we compare the accuracy of three different algorithms: k-nearest neighbors, logistic regression and decision trees by using 230 observations in the training set and 40 in the test set. The accuracy scores for the selected algorithms are very similar and are more or less 85%. After plotting the ROC curves and calculating the area under the curve, we can finally conclude that the logistic model is the most accurate for our study set.

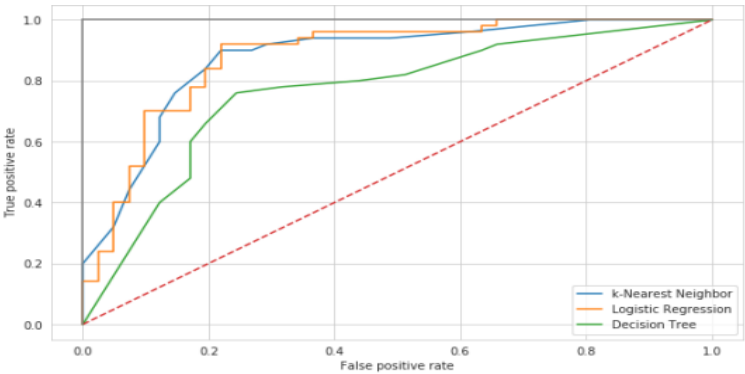


Figure 4. Comparison of ROC curves

4. Conclusions

Since the monitoring of hospitalized patients is currently being carried out only in critical cases, especially in intensive care units, using inflexible and costly systems, it has been considered useful to the medical world to demonstrate the construction of a system that highlights the technological advance in recent years in the field of monitoring the heart health, bringing benefits in terms of cost, flexibility, interoperability and usability. In order to meet these factors, it was chosen to monitor the heart's activity redundantly to enhance patients' comfort, provide a reliable set of data and building an accessible client application from mobile devices for quasi-real-time viewing of the patients' situation. The doctors and caregivers monitor the patient in real time using the data received through the server. Ensuring the interoperability between the mobile application and other medical units using FHIR, as in our particular system is presented the communication with an Obstetrics-Gynecology Department, improves the patient health and the doctor has a better support for a better treatment, due to continuity of healthcare records.

References

- [1] Heart Disease and Stroke Statistics—2016 Update
- [2] A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee.
- [3] E.Jung, J. Kim, KY Chung, D. K. Park, Mobile healthcare application with EMR interoperability for diabetes patients, *Cluster Computing*, 17 (3)(2014), 871 – 880.
- [4] J.Coons, R. Patel, K. Coley, P. Empey, Design and testing of Medivate, a mobile app to achieve medication list portability via Fast Healthcare Interoperability Resources, *Journal of the American Pharmacists Association*, 59 (2) (2019), S78 – S85.
- [5] .F Alves, L. Ferreira, N. Lopes, M.L.R. Valera, H. Castro, P.S. Avila, H.A. Teixeira, G.D. Putnik, M.M. Cruz-Cunha, FHIRbox, a cloud integration system for clinical observations, *Procedia Computer Science* 138 (2018), 303 – 309.
- [6] H7 FHIR (Fast Healthcare Interoperability Resources), www.hl7.org/fhir, Accessed in 15.08.2019.
- [7] Azure API for FHIR, <https://docs.microsoft.com/en-us/azure/healthcare-apis/>, Accessed in 20.08.2019.
- [8] M. Crisan-Vida, A. Mihaltan, L. Stoicu-Tivadar, A. Moza, E. Bernad, ICT Supporting Treatment for High Risk Situations in Obstetrics, The 6th IEEE International Conference on E-Health and Bioengineering - EHB 2017.

Data Fusion to Convert Drug Consumption Quantities into Defined Daily Doses

Adrien UGON^{a,1}, Sabine GUESSANT^b, and Rosy TSOPRA^{c,d,e}

^aESIEE-Paris, Cité Descartes BP99, Noisy-le-Grand, 93162, France.

^bDepartment of Public Health, Hôpital Saint-Antoine, Assistance Publique-Hôpitaux de Paris, Paris, France.

^cINSERM, Centre de Recherche des Cordeliers, UMRS 1138, Université Paris-Descartes, Université Sorbonne Paris Cité, France.

^dDepartment of Medical Informatics, Hôpital Européen Georges Pompidou, AP-HP, Paris, France

^eLTSI Université Rennes 1 Team Health Big Data, Rennes

Abstract. Analysis of consumptions has proved a misuse of antibiotics, despite the existence of national requirements. To be able to compute drugs consumption quantities of highly heterogeneous drugs expressed in various doses units, the World Health Organization has defined a defined daily dose. A methodology has been also defined from previous work to compute manually the drugs consumptions in daily defined dose. We automated this methodology by using data fusion on data retrieved from different sources including a French public database and the World Health Organization website. Evaluation proved the efficiency of the approach, except for inconsistency cases. We identified these cases and proposed a solution to avoid them.

Keywords. Drug database, Health Information Interoperability, Automatic data processing, Drug misuse

1. Introduction

French authorities in public health study and publish regularly results about antibiotics consumption in France and in the world [1]. More generally, drug consumption trends are observed [2]. Due to a misuse of antibiotics, it is thought that “Antimicrobial resistance could become one of the leading causes of death in the world” [3]. To obtain these results it is necessary to compute automatically measures reflecting the consumption of antibiotics. World Health Organization (WHO) has defined a unit to express to consumptions, called daily defined dose (DDD). Drug Defined Dose is defined by WHO as “The assumed average maintenance dose per day for a drug used for its main indication in adults.” [4]. It was first introduced by S. Natsch et al [5]. It was observed that, despite the existence of national requirements, hospital antibiotic stewardship programs are not fully implemented in France [6] Hence, it is necessary to automate the computation of the drugs consumption, and more precisely antibiotics consumption, in DDD using data and information stored in the clinical information systems.

¹ Corresponding Author, E-mail: adrien.ugon@esiee.fr.

The ATC (Anatomical Therapeutic Chemical) Classification system hierarchically classifies therapeutic drugs into 5-level groups, considering the organ or system on which they act and their chemical, pharmacological and therapeutic properties. Each group is identified by a code.

WHO provides a DDD, for each ATC code, and administration route, and sometimes pharmaceutical form, as a dose expressed with a value and a unit. They can be retrieved from a website using text mining methods [7]. Hence, the computation of drug consumption quantity can be considered as a unit conversion. Methodology to measure antibiotic consumption has been proposed in previous work [8] but needs to be automated.

Drug information is provided by French public database [9] under heterogeneous formats. Structured information are provided in text files following the tabula separated values format. Information on specialties is given in a first file called *CIS_bdpm.txt*. Each specialty is identified by a CIS (French for “Specialty Identifier Code”) code, and described with the name, the administration route, or the pharmaceutical form and other administrative information about the specialty. Each specialty can be sold in different presentations, each being identified by a CIP (French for “Presentation Identifier Code”). All presentations are provided in a file called *CIS_CIP_bdpm.txt*, providing CIP code, CIS code, presentation name, price and other administrative information about presentation. A third file, called *CIS_COMPO_bdpm.txt* provides information about the substances that are part of a medicinal product, among which the name of the substance, the dose, the nature (active substance or therapeutic fraction). In addition of this information, HTML pages provide other textual information that can be parsed with text mining method. The ATC Code needs to be retrieved from these HTML pages.

This paper aims at designing a tool to automate the computation of the drugs consumption in DDD using information provided by a French public drug database. Section 2 will describe the methods. Results will be provided in section 3. Finally a discussion will be proposed in section 4.

2. Methods

Our approach is based on data fusion. It has been proved that symbolic approaches are efficient to fuse heterogeneous, and eventually multimodal data coming from various scattered sources [10]. The system is composed of two consecutive steps: (1) data retrieval (2) data fusion.

2.1. Data retrieval

To be able to compute the consumptions in DDD, it is necessary to retrieve all necessary data. In the clinical information systems, consumed drugs are identified by their CIP code, which is given in the *CIS_CIP_bdpm.txt* file. Then, we need to know the dose of all substances that compose the drug, which is provided in the *CIS_COMPO_bdpm.txt* file. The link between the two files is the CIS, identifying a specialty, provided in the two files. With this information, from these two files, it is thus possible to retrieve the dose of all substances that compose a drug.

To retrieve the DDD defined by the WHO, it is necessary to have the administration route, which is provided in the *CIS_bdpm.txt* file given the CIS Code. The ATC Code can be retrieved from the HTML page, given a CIS Code and using a regular expression. Because of errors in the ATC Code – some of them used the “O” character instead of the

“0” number – it was necessary to use an adapted regular expression to find all ATC codes for a given CIS code. Hence, we used the following regular expression: “[A-Za-z][0-9O]{2}[A-Za-z]{2}[0-9O]{2}”. It is then possible to retrieve the DDD value and units by parsing the ATC/DDD Web page for a given ATC Code.

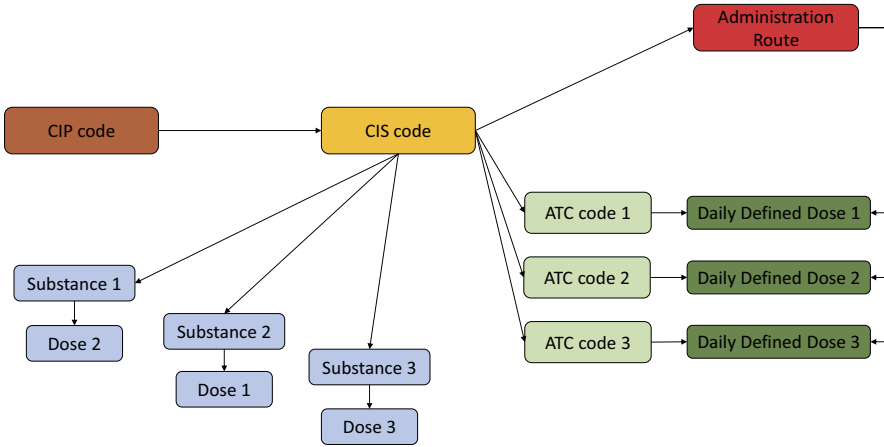


Figure 1. Data dependencies

Figure 1 shows the relationships between the different required data. In a yellow rectangle is represented the CIS code, which is the key data allowing to link all other data. In a brown rectangle is represented the data provided in the *CIS_CIP_bdpm.txt* file; in blue rectangles are represented data provided by *CIS_COMPO_bdpm.txt* file; in red rectangle is represented the data provided by *CIS_bdpm.txt* file; in pale green rectangle is represented the data provided by HTML pages of the public drug database; in forest green is represented the data provided in the WHO ATC/DDD Webpage.

2.2. Data fusion

Respecting the principles of high-abstraction-level symbolic fusion [11], all data is represented as a list of multimodal data, indexed by an identifying code. Fusion strategies are defined to fuse data considering the relationships pointed out in Figure 1.

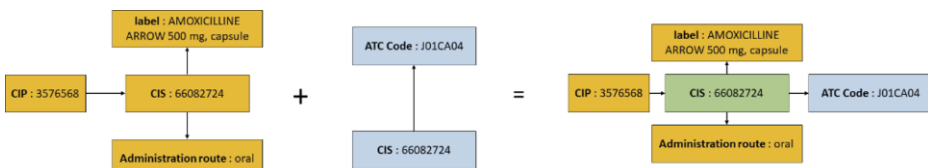


Figure 2. High-abstraction-level fusion strategies

Figure 2 shows how high-abstraction-level fusion strategies are defined considering the data model presented in Figure 1. Multimodal data is represented by conceptual

graphs allowing the fusion. By applying fusion strategies, we can enrich information about a medicinal product identified by a CIP code.

Once we have got the DDD provided by the WHO and the dose of all substances of a drug, we are able to compute their value expressed in DDD by dividing each substance dose by each DDD associated with a specialty. As a consequence, one value will be computed for each substance of the drug.

3. Results

The method was applied to the 20 479 drugs, identified by a different CIP code, from the *CIS_CIP_bdpm.txt* file, present in the French public drug database; we were able to compute the dose expressed in DDD for 9 264 of them (45.23 %). We were unable to compute the DDD for 249 drugs for inconsistency reasons which will be discussed in the next section. The main encountered problem was that for 11 027 drugs (53.8%), the DDD was not defined by WHO.

By focusing on antibiotics, we found 966 entities, among which we were able to compute the dose expressed in DDD for 780 (80.8%) of them.

4. Discussion

Several inconsistencies were observed in data and information provided by the French drug public database, making sometimes impossible to compute the DDD.

First, we observed missing data. Although information was provided for a CIP code, information about corresponding CIS code was missing, either about the composition of the drug, or about the administration route. This should be solved by checking the completeness of the information provided in the database.

Second, only one ATC Code is given – at most – on the web page of the French public drug database, which does not reflect some cases where a drug can have several ATC codes. This is the case, for instance, of the prednisolone, which has seven ATC codes; the only that can be retrieved from the web page is H02AB06. This can be solved by adding all the ATC codes to the structured files.

Third, as already mentioned, there can be errors in the ATC code, which can be missing, or can be shorter than expected, or containing spaces and containing the “O” character instead of a “0” number. For these reasons, text mining methods using a regular expression are not efficient to retrieve all ATC codes. This should be solved by providing a clean value of the ATC codes in the structured files.

Fourth, there is no alignment of administration routes representation in the drug database and in the WHO ATC/DDD webpage. To be able to compute the consumption in DDD, it is necessary, in a preliminary step, to align manually the different values used in the two sources. Standardizing the expression of administration routes would allow to prevent this issue.

Fifth, doses are sometimes expressed under a form that is impossible to process automatically. For instance, doses like “2ch to 30ch and 4dh to 60dh”, or “75 mg equivalent to 69,6mg of risedronic acid” cannot be processed automatically.

Sixth, sometimes substance doses and WHO DDD are expressed in incompatible units : micrograms and Thousand Units, milliliters and milligrams. Constraints on the

dose data format, combining a numeric value to a unit would allow to avoid the two last presented issues.

Finally, there exists some ATC codes without any DDD by the WHO. For instance, “J01DA33”, “G04CA05” or “R02AA19” have no DDD by WHO.

Acknowledgement

Funding was obtained from the Agence Nationale de Sécurité du Médicament et des Produits de santé (ANSM). APP 2016—RaMiPA Project (Raisonnement pour Mieux Prescrire les Antibiotiques [Reasoning for a better antibiotic prescription]).

References

- [1] C. Jackson, Y. Hsia, J.A. Bielicki, S. Ellis, P. Stephens, I.C.K. Wong, M. Sharland, Estimating global trends in total and childhood antibiotic consumption, 2011-2015, *BMJ Global Health* **4** (2019), e001241.
- [2] Y. Hsia, M. Sharland, C. Jackson, I.C.K. Wong, N. Magrini, J.A. Bielicki, Consumption of oral antibiotic formulations for young children according to the WHO Access, Watch, Reserve antibiotic groups: an analysis of sales data from 70 middle-income and high-income countries, *Lancet Infect Dis* (2019) 19(1):67-75
- [3] <https://www.santepubliquefrance.fr/les-actualites/2018/consommation-d-antibiotiques-et-antibioresistance-en-france-en-2017>
- [4] https://www.who.int/medicines/regulation/medicines-safety/toolkit_ddd/en/
- [5] S. Natsch, Y.A. Hekster, R. de Jong, et al. Application of the ATC/DDD methodology to monitor antibiotic use, *European Journal of Clinical Microbiology & Infectious Diseases* **17** (1998); 20-24
- [6] F. Binda, G. Tebano, M.C. Kallen, J. ten Oever, M.E. Hulscher, J.A. Schouten, C. Pulcini, Nationwide survey of hospital antibiotic stewardship programs in France, *Médecine et Maladies Infectieuses*, (2019)
- [7] https://www.whocc.no/atc_ddd_index/
- [8] J.M. Hutchinson, D.M. Patrick, F. Marra, H. Ng, W.R. Bowie, L. Heule, M. Muscat, D.L. Monnet, Measurement of antibiotic consumption: A practical guide to the use of the Anatomical Therapeutic Chemical classification and Defined Daily Dose system methodology in Canada, *The Canadian journal of infectious diseases* 15(1) (2004), doi:10.1155/2004/389092
- [9] <http://base-donnees-publique.medicaments.gouv.fr/>
- [10] A. Ugon, C. Philipe, A. Kotti, M.-A. Dalloz, A. Pinna, Automatic sleep stages classification combining semantic representation and dynamic expert system 264 (2019), 848-852
- [11] C. Laudy, J.-G. Ganascia, Introducing Semantic Knowledge in High-Level Fusion, The SIMA workshop of Milcom2009, Boston, MA, United States (Oct 2009), 1-7, 10.1109/MILCOM.2009.5379903

Data Integration into OMOP CDM for Heterogeneous Clinical Data Collections via HL7 FHIR Bundles and XSLT

Patrick FISCHER^{1, a}, Mark R. STÖHR^b, Henning GALL^b,
Achim MICHEL-BACKOFEN^c and Raphael W. MAJEED^b

^a*Institute of Medical Informatics, Faculty of Medicine, Justus-Liebig-University, Giessen, Germany*

^b*UGMLC, German Center for Lung Research (DZL), Justus-Liebig-University, Giessen, Germany*

^c*Data-Integration-Center, Faculty of Medicine, Justus-Liebig-University, Giessen, Germany*

Abstract. Data integration is an important task in medical informatics and highly impacts the gain out of existing health information data. These tasks are using implemented as extract transform and load processes. By introducing HL7 FHIR as an intermediate format, our aim was to integrate heterogeneous data from a German pulmonary hypertension registry into an OMOP Common Data Model. First, domain knowledge experts defined a common parameter set, which was subsequently mapped to standardized terminologies like LOINC or SNOMED-CT. Data was extracted as HL7 FHIR Bundle to be transformed to OMOP CDM by using XSLT. We successfully transformed the majority of data elements to the OMOP CDM in a feasible time.

Keywords. ETL, OMOP, HL7 FHIR, LOINC, SNOMED-CT

1. Introduction

International collaborative research faces a variety of challenges. One includes integrating data from different heterogeneous data sources into a data repository or data registry. Such registries can serve a lot of different purposes such as drug safety analysis [1], longitudinal analysis of certain cohorts [2] or as in our, and many other cases [3] a disease specific registry to promote research in that field. As part of the Pulmonary Vascular Research Institute, a collaboration of healthcare professionals and researchers working in the field of pulmonary hypertension (PH), an international PH registry/data repository needs to be established.

We chose the OHDSI OMOP Common Data Model [1] to be the foundation of this registry. The OMOP CDM is a database specification, based on standardized vocabularies like SNOMED-CT, LOINC or RxNorm, with the aim to provide an analytic-friendly non-complex representation of medical data [1].

To populate a registry with data from existing databases, usually, Extract-Transform-Load tools (ETL) are used to transform and transfer data to the target schema. Popular tools in this regard are Talend Open Studio or Pentaho, which are used to define the ETL process as a dataflow from source to target system. However, by using such software a complete ETL process needs to be implemented for each source separately. This is also the recommended way of designing ETL processes for the OMOP CDM. Because we want to connect multiple heterogeneous databases, a common parameter set was predefined. We plan to use HL7 FHIR [4] as an intermediate format for the local data extraction. The FHIR resource will be fed into an XSLT processor to generate CSV files in the OMOP schema. Similar international and national approaches using intermediate formats or HL7 FHIR together with OMOP CDM have led to good results [5–7].

Therefore, our aim is to evaluate the feasibility of HL7 FHIR Bundle and XSLT as a generic ETL process to populate an OMOP CDM. Feasibility will be assessed by the computation time and coverage of source data in the target CDM.

2. Method

2.1. Common data set

To integrate data from multiple registries into a single data warehouse, a common data set is needed. A meeting of domain experts from two large PH registries from both the UK and Germany was scheduled to discuss all data items and produce a common data set definition. In a second step, this data set is annotated with international standard terminology (LOINC, SNOMED-CT, etc.) by documentation experts.

2.2. Source data

As source data, we used a German PH registry database. Its data is stored in a Microsoft Access database with 22 proprietary tables and 18 user forms. Additionally, two EXCEL sheets are used to collect data about specimen and echocardiography. All tables are exported to CSV format using MS Access’s standard export functionality. For ETL-preparation, we used HStream-import to convert the proprietary CSV files to a standard FHIR Bundle collection. For this purpose, an XML configuration file is written which includes mapping of source field names to standard terminology as defined by the common data set. The resulting HL7 FHIR Bundle resource contains Patient, Encounter and Observation resources and conforms to the latest FHIR standard 4.1.0 [4]. An example of the Bundle structure containing the source data is depicted in Figure 1.

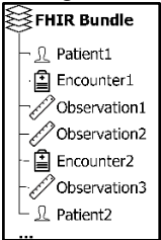


Figure 1. Example FHIR Bundle structure of provided input data

2.3. Mapping to OMOP domains

The OMOP CDM splits data into different domains depending on their scope. Condition, Drug, Procedure, Measurement are common medical domains, which are mapped to separate tables in the OMOP CDM. We used Athena [8] to browse the complete OMOP vocabularies to find the correct target table for the data elements of our defined data set. We did so by searching for the respective SNOMED-CT, LOINC or ATC code and retrieving the OMOP domain of the resulting concepts, as well as the concept ids.

2.4. Extending OMOP vocabularies

There are many subtypes of pulmonary hypertension depending on the etiology or comorbidities of the patient. Those subtypes are classified using the clinical classification of the World Symposium on Pulmonary Hypertension of 2013, which is also referred to as PH Nice Classification [9]. Since the Nice classification is not part of the standardized vocabularies already included to OMOP, we had to add it manually.

By using the hierarchical classification and the ER model of the OMOP standardized vocabularies [10], we generated CSV files to bulk load the Nice classification to OMOP.

2.5. Data transformation using XSLT and XPath

To transform the source data, provided as HL7 FHIR Bundle, to the OMOP schema we used XSLT (v2.0) [11] and XPath (v2.0) [12] to generate CSV for each used OMOP target table. The resulting CSV files contain all required information for the respective table, e.g. Measurement or Patient.

3. Results

3.1. Mapping to standard terminology

The predefined common dataset consists of 141 distinct data elements all represented by a specific code. Out of the 141 elements, 36 have been mapped to SNOMED-CT, 38 to LOINC, 27 to ATC and one to ICD-10. 27 of the remaining 39 data elements belong to the PH Nice classification, while no standard code was found for 12 parameters.

3.2. Algorithm

The transformation is split up into different files, one for each table that will be populated. Currently our approach populates the following OMOP tables with all required fields: *PERSON*, *OBSERVATION_PERIOD*, *CONDITION_OCCURRENCE*, *DRUG_EXPOSURE*, *PROCEDURE_OCCURRENCE*, *OBSERVATION* and *VISIT_OCCURRENCE*

The transformation structure is in principle the same for each table. A generic XSLT template is defined, to generate one row for each entry to the respective table. This generic template also extracts metadata about the entry like the date time or references to patients or encounters. Additionally, we defined parameter-specific templates, which map one parameter to its correct concept id in OMOP. Those parameter-specific

templates call the generic template with the correct parameter specific data like concept id, status ids or type ids for the respective parameter.

The resulting CSV files are then bulk loaded to the database. A schematic illustration of the ETL process is shown in Figure 2.

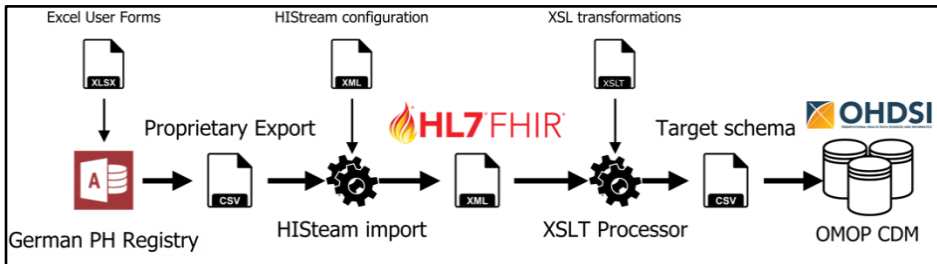


Figure 2. Schematic illustration of ETL process, showing involved software, including configurations and temporary result files

3.3. Transformation results and performance

Except for the 12 data elements of the common data set, where no standard concept in the standardized vocabularies could be found, every element was successfully mapped to the OMOP domains resulting in an overall coverage of about 91.5 %. We applied for new codes within LOINC for five data elements [13]. Additionally, we added 34 new concepts, 54 concept relationships and 68 concept ancestor entries to the vocabularies and therefore achieved 100 % coverage of the PH Nice classification in OMOP.

In total, data of 3,887 patients was transformed to OMOP in 15 minutes and 29 seconds using Saxon-HE v9.9.1 XSLT processor (on Intel Core i7-8550U with 256 GB SSD). To populate all required fields, we assigned default values for the parameters, which were not available in the source data. Those fields were for example: *visit_type*, *admitted_from*, *discharged_to* and others. In addition, fields like ethnicity and race have been set to default values for the first prototype, but those will be correctly mapped in the future as well.

4. Discussion

Based on the achieved results we conclude the feasibility for HL7 FHIR Bundle and XSLT as part of a generic ETL process. The majority of data elements were successfully transformed to the target scheme in reasonable time. Additionally, the customization of the available ETL process to connect new source databases reduces to only adjusting the data extraction process and HISstream configuration to provide data as HL7 FHIR Bundle. Therefore, providing a HL7 FHIR export for a local data source can increase interoperability of the included data, not only for data warehousing, but also for communication with other systems.

Limitations of this work include the necessary synchronization between HISstream configuration and XSL transformations in order to transform all available data elements to OMOP. Additionally, the mapping of standard concepts is currently hard coded, by leveraging the rich vocabularies provided as CSV the mapping can be designed in a more

generic way. We also plan to reduce the computational time, even though 15 minutes for the transformation is already feasible because data will only be loaded periodically.

References

- [1] J.M. Overhage, P.B. Ryan, C.G. Reich, A.G. Hartzema, and P.E. Stang, Validation of a common data model for active safety surveillance research, *J. Am. Med. Inform. Assoc.* **19** (2012) 54–60. doi:10.1136/amiainl-2011-000376.
- [2] M. Garza, G. Del Fiore, J. Tenenbaum, A. Walden, and M.N. Zozos, Evaluating common data models for use with a longitudinal community registry, *J. Biomed. Inform.* **64** (2016) 333–341. doi:10.1016/j.jbi.2016.10.016.
- [3] M. Coleman, D. Forman, H. Bryant, J. Butler, B. Rachet, C. Maringe, U. Nur, E. Tracey, M. Coory, J. Hatcher, C. McGahan, D. Turner, L. Marrett, M. Gjerstorff, T. Johannesen, J. Adolfsson, M. Lambe, G. Lawrence, D. Meechan, E. Morris, R. Middleton, J. Steward, and M. Richards, Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data, *The Lancet*. **377** (2011) 127–138. doi:10.1016/S0140-6736(10)62231-3.
- [4] HL7 International, FHIR v4.1.0, *FHIR*. (2018). <http://build.fhir.org/> (accessed September 30, 2019).
- [5] C. Maier, L. Lang, H. Storf, P. Vormstein, R. Bieber, J. Bernarding, T. Hermann, C. Haverkamp, P. Horki, J. Laufer, F. Berger, G. Höning, H.W. Fritsch, J. Schüttler, T. Ganslandt, H.U. Prokosch, and M. Sedlmayr, Towards Implementation of OMOP in a German University Hospital Consortium, *Appl. Clin. Inform.* **09** (2018) 054–061. doi:10.1055/s-0037-1617452.
- [6] J. Guoqian, K. Richard, P. Eric, and S.H. R, Building Interoperable FHIR-Based Vocabulary Mapping Services: A Case Study of OHDSI Vocabularies and Mappings, *Stud. Health Technol. Inform.* (2017) 1327–1327. doi:10.3233/978-1-61499-830-3-1327.
- [7] J. Guoqian, K.R. C, S.D. K, P. Eric, and S.H. R, A Consensus-Based Approach for Harmonizing the OHDSI Common Data Model with HL7 FHIR, *Stud. Health Technol. Inform.* (2017) 887–891. doi:10.3233/978-1-61499-830-3-887.
- [8] OHDSI, Athena, (2019). <http://athena.ohdsi.org/search-terms/terms> (accessed September 25, 2019).
- [9] G. Simonneau, M.A. Gatzoulis, I. Adatia, D. Celermajer, C. Denton, A. Ghofrani, M.A. Gomez Sanchez, R. Krishna Kumar, M. Landzberg, R.F. Machado, H. Olschewski, I.M. Robbins, and R. Souza, Updated Clinical Classification of Pulmonary Hypertension, *J. Am. Coll. Cardiol.* **62** (2013) D34–D41. doi:10.1016/j.jacc.2013.10.029.
- [10] OHDSI, OHDSI/CommonDataModel, *GitHub*. (2019). <https://github.com/OHDSI/CommonDataModel> (accessed September 25, 2019).
- [11] XSL Transformations (XSLT) Version 2.0, (n.d.). <https://www.w3.org/TR/xslt20/> (accessed January 7, 2020).
- [12] XML Path Language (XPath) 2.0 (Second Edition), (n.d.). <https://www.w3.org/TR/xpath20/> (accessed January 7, 2020).
- [13] LOINC, Submitting New Term Requests, *LOINC*. (2019). <https://loinc.org/submissions/new-terms/> (accessed September 25, 2019).

Data Quality Challenges in a Learning Health System

Michail SARAFIDIS ^a, Marilena TAROUSI ^a, Athanasios ANASTASIOU ^{a,1}, Stavros PITOGLOU ^b, Efstratios LAMPOUKAS ^b, Athanasios SPETSARIAS ^b, George MATSOPOULOS ^c and Dimitrios KOUTSOURIS ^a

^a*Institute of Communication and Computer Systems (ICCS), Athens, Greece*

^b*Research & Development Dpt., Computer Solutions SA, Athens, Greece*

^c*School of Electrical & Computer Engineering, National Technical University of Athens, Athens, Greece*

Abstract. This paper discusses the topic of data quality, which concerns the global research and business community and constitutes a challenging task. The data quality prerequisite becomes even more critical when it pertains to critical and sensitive data, such as the healthcare domain data. To begin with, the paper outlines the basic definitions and concepts of data quality and its dimensions. The related research work on data quality assessment is presented and our approach for data quality assurance is introduced. This approach is implemented in our designed cloud platform, called MODELHealth, which is intended for supporting clinical work and administrative decision-making process.

Keywords. data quality, health data, ehr, quality assurance

1. Introduction

Today's world is entirely data-driven. Enormous amounts of data are generated at every point of interaction. Therefore, how this data can be used effectively is a matter of great importance. For this data to be utilizable, they have to meet some requirements and be reliable, up-to-date, and of high quality. Thus, data quality is one of the biggest challenges in the digital world. The high volume and complexity of data collected across multiple sources create an enormous task in terms of data quality management. It is essential to mention that the issues in data quality can directly affect analytics and decision-making.

For these reasons, data quality is recognized as widely considered to be a critical factor, especially in the healthcare sector. This paper outlines the basic definitions and concepts of data quality and proposes a new approach for data quality assurance.

¹ Athanasios Anastasiou, Institute of Communication and Computer Systems (ICCS), Zografou Campus, 9 Iroon Polytechniou Str., 15780, Athens, Greece; E-mail: aanastasiou@biomed.ntua.gr.

2. Quality of Health data

2.1. Data Quality Definition and dimensions

Data are real-world objects, with the property of storing, retrieving and elaborating through an algorithmic process and can be communicated between two entities. **Data quality** has a distinct definition in various fields and periods. Multiple definitions of data quality have been proposed, but data is generally considered high quality if **they are fit “for intended uses in operations, decision making, and planning”** [1] or, more generally, **“for the purposes data consumers want to apply them”** [2]. Indeed, the quality of data is critical for improvement process activities, and it can be addressed in various fields inclusive of finance, statistics, computer science, and medicine.

The primary phase in any data quality process is to define the properties of data quality; those properties are defined as data quality dimensions (DQDs). A **Data Quality Dimension** is a term used to define a data quality measure that can describe numerous data elements including attribute, record, table, system or more abstract groupings, such as business unit, company or product range [3]. Researchers have recognized various numbers of dimensions for data quality, but the most important are the six following core dimensions: Accuracy, Completeness, Consistency, Timeliness, Uniqueness and Validity.

Accuracy of the data is the degree to which data is correct, reliable and certified. In other words, data are accurate in case of data values stored in the database are equivalent to real-world values. It is a measure of the correctness of the content of the data.

Completeness of the data concerns the extent to which values appear in a dataset. With regard to a separate datum, only two conditions are potential: The attribute under consideration has a value or not.

Suraj Juddoo et al. suggested that the two common most significant DQDs to consider in the context of big datasets for the health domain are accuracy and completeness [4].

Consistency of the data is the extent to which information is presented in the identical format and compatible with previous data.

Timeliness of data regards only the interval between an alteration of a real-world state and the outcoming modification of the information system state. Timeliness has two constituents: age and volatility, which are a measure of how old the information is and a measure of information instability, respectively.

Uniqueness of data determines the extent to which there are no duplicate values.

Validity of data refers to data that have been collected in conformity with any rules or definitions that apply to that data. This will enable benchmarking between organizations and over time.

According to the *“Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data”* [5], an initiative aimed at providing a solution to the inconsistency of data quality terminology used throughout the relevant scientific literature, the main proposed categories of healthcare data quality are:

Conformance: The data should be compliant as far as format, relational reference, and computational accuracy are concerned. The subcategories of conformance are:

- **Value Conformance:** Values and format of the data should adhere to all given prespecified data architecture constraints.

- **Relational Conformance:** Data should be consistent with the given structural constraints (nullability, referential integrity, etc.).
- **Computational Conformance:** If there are computed values derived from existing data, they should be correct and consistent between computations based on the same specifications (internal or external programs).

Completeness: In accordance with the respective dimension mentioned above, data should be complete, i.e., conforming to the expected frequencies of every data attribute value presence. In this context, completeness is not concerned with the data values, structure or plausibility.

Plausibility: The data values should be believable/truthful when put in context (other variable values, temporal sequences, state transitions etc.). Plausibility subcategories are:

- **Uniqueness Plausibility:** Objects represented by data should not appear multiple times where duplication is not explicitly justified.
- **Atemporal Plausibility:** Observed data values and/or distributions should agree with domain knowledge, trusted external sources or established “gold standards”.
- **Temporal Plausibility:** Time-dependent variable values should change as expected, taking under consideration established temporal properties.

2.2. Importance of Data Quality in Electronic Health Record

E-health constitutes a recognized term and is close related to the usage of **Electronic Health Records (EHRs)**, so as to achieve adequate communication between healthcare sectors. Specifically, various sources of healthcare data exist, such as the internet, electronic patient records, data analysis tools, communication between health experts and patients, electronic health devices. Data management and analysis tools aim to improve healthcare organizations operation in terms of obtaining useful information and knowledge via conversion of the aforementioned vast resources [6].

It is crucial that the collected data are precise, comprehensive, reliable, comprehensible and available to all stakeholders such as patients, health experts, legal authorities, and government authorities [7]. Therefore, data quality issue constitutes a significant challenge concerning health records and is able to provoke pivotal effects on every health sector operation. A considerable amount of lethal medical errors concerns the result of missing or incomplete data about medication, instructions, and treatment plans [8].

In addition, adequate data quality plays a significant role in the successful implementation of machine learning methods, and it presents one of the significant challenges, especially in the healthcare domain [9], [10]. Poor quality input data are, most of the times, a critical liability in the machine learning process as, in essence, they carry an altered state of the underlying patterns that represent the “truth” of the data, leading to an altered “reality” perceived by the algorithm. It’s self-evident that regarding a learning process, if the foundation -the learning material- is critically impaired, one should not have many expectations of the outcome.

The most common source of data used in machine learning processes of medical significance is operational software systems in clinical environments, mainly EHR transactions. This secondary analysis setting, meaning the use of the data for research analytics, which is not their original purpose of collection, has been researched

extensively towards assessing and dealing with the factors that affect data quality [11]–[13]. It has been found that *“EHR data from clinical settings may be inaccurate, incomplete, transformed in ways that undermine their meaning, unrecoverable for research, of unknown provenance, of insufficient granularity, and incompatible with research protocols”* [11].

3. Health Data Quality in MODELHealth

MODELHealth is a cloud platform that facilitates the application of Machine Learning methods, towards processing health data, in order to support clinical work or/and the administrative decision-making process, in all levels of health services provision. It is a “holistic” approach to the subject of developing and exploiting Machine Learning algorithms. It covers the whole lifecycle of the respective process, from querying, homogenizing, anonymizing and enriching raw data, to the final provision of the resulting algorithms via Application Program Interfaces, in order for them to be consumed by any authorized Information System. For projects like MODELHealth, data retrieval is a critical step, as the raw data that are produced by querying a real-world database need to be handled on various levels for a number of reasons before they end up on the central data repository, in order to participate in the training and validation of machine learning algorithms.

To this end, the ETL (Extract, Transform & Load) mechanism of the MODELHealth platform contains a **dedicated Quality Assurance (QA) module**, as well as an Entity Mapping module that handles ontologies and central database homogeneity and an anonymization module that handles patient privacy and regulatory compliance. Both latter modules lie outside the scope of this text.

Regarding the data quality Assessment, the QA module of the ETL system is capable of performing a number of tests on the incoming data, based on a customizable and extendable rule-set that quantifies data quality in the form of test success-failure percentages. There will be rules covering the whole aforementioned spectrum of healthcare data quality: atemporal completeness, value conformance, and atemporal plausibility. These are the factors that are primarily investigated through data quality checks by the organizations which are currently engaged in data quality assessment [14]. Via centrally defined customizable thresholds for each active rule, the administrators can set the quality tolerance of the data retrieval process, thus controlling the minimum quality requirements of the data that actually make it to the central data repository. This is accomplished by “instructing” the decentralized ETL mechanisms to reject data that do not meet the desired criteria.

More specifically, in order to pursue the quality standards, the developed QA module will be based on the research conducted by Nicole G. Weiskopf et al. [15]. The tool’s core framework takes into account three different forms of data: complete, correct, and current data. Data quality assessment issues diversify from one another forms of data. Furthermore, the operationalization of the above forms varies depending on the three main types of HER data: patients, variables, and time. As a result, the combination of the aforementioned 3x3 elements produces nine constructs that lead to different methodological recommendations, in other words guideline-based best practices, for EHR data quality assessment.

4. Conclusions

This paper has described a spectrum of issues associated with the quality of health-care data. Because contemporary EHR systems suffer from many shortcomings, the medical data that they produce are also often flawed. Data-quality problems can compromise the value of databases for scientific research, quality assessment, public health, and other purposes. Analytical insights in healthcare should always be probed for data quality problems. Also, the value derived from the use of analytics should dictate the requirements of data quality. This is emphasized because most analytical tools assume that the data are of very high quality. Based on this premise, MODELHealth attempts to apply contemporary data quality checks, towards a systematic approach to data quality, in order to ensure a reliable and viable developed asset for healthcare organizations for the holistic implementation of machine learning processes.

Acknowledgment

This research has been co-funded by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under call Research-Create-Innovate (Project Code: T1EDK-04066).

References

- [1] T. C. Redman and D. Ph, *Data Driven : Profiting from Your Most Important Business Asset*. 2008.
- [2] Data International, *DAMA-DMBOK : data management body of knowledge*, 2nd ed. Technics Publications, 2017.
- [3] F. Sidi, P. H. Shariat Panahy, L. S. Affendey, M. A. Jabar, H. Ibrahim, and A. Mustapha, 'Data quality: A survey of data quality dimensions', 2012 Int. Conf. Inf. Retr. Knowl. Manag., pp. 300–304, Mar. 2012.
- [4] S. Juddoo and C. George, 'Discovering Most Important Data Quality Dimensions Using Latent Semantic Analysis', 2018 Int. Conf. Adv. Big Data, Comput. Data Commun. Syst., pp. 1–6, 2018.
- [5] M. G. Kahn et al., 'A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data', *eGEMs (Generating Evid. Methods to Improv. patient outcomes)*, vol. 4, no. 1, p. 18, 2016.
- [6] Y. Sun, T. Lu, and N. Gu, 'A method of electronic health data quality assessment: Enabling data provenance', *Proc. 2017 IEEE 21st Int. Conf. Comput. Support. Coop. Work Des.* pp. 233–238, 2017.
- [7] M. Botha, A. Botha, and M. Herselman, 'Data quality challenges: A content analysis in the e-health domain', 2014 4th World Congr. Inf. Commun. Technol. WICT 2014, pp. 107–112, 2014.
- [8] R. L. Leitheiser, 'Data quality in health care data warehouse environments', *Proc. Hawaii Int. Conf. Syst. Sci.*, vol. 00, no. c, p. 152, 2001.
- [9] S. Pitoglou, 'Machine Learning in Healthcare, Introduction and Real World Application Considerations', *Int. J. Reliab. Qual. E-Healthcare*, vol. 7, no. 2, pp. 27–36, Apr. 2018.
- [10] V. N. Gudivada, A. Apon, and J. Ding, 'Data Quality Considerations for Big Data and Machine Learning : Going Beyond Data Cleaning and Transformations', vol. 10, no. 1, pp. 1–20, 2017.
- [11] W. R. Hersh et al., 'Caveats for the use of operational electronic health record data in comparative effectiveness research', *Med. Care*, vol. 51, no. 8 Suppl 3, pp. S30-7, Aug. 2013.
- [12] J. S. Brown, M. Kahn, and S. Toh, 'Data quality assessment for comparative effectiveness research in distributed data networks', *Med. Care*, vol. 51, no. 8 Suppl 3, pp. S22-9, Aug. 2013.
- [13] A. Rusanov, N. G. Weiskopf, S. Wang, and C. Weng, 'Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research', *BMC Med. Inform. Decis. Mak.*, vol. 14, no. 1, p. 51, Dec. 2014.
- [14] T. J. Callahan et al., 'A Comparison of Data Quality Assessment Checks in Six Data Sharing Networks', *eGEMs (Generating Evid. Methods to Improv. patient outcomes)*, vol. 5, no. 1, p. 8, Jun. 2017.
- [15] N. G. Weiskopf, S. Bakken, G. Hripesak, and C. Weng, 'A Data Quality Assessment Guideline for Electronic Health Record Data Reuse', *eGEMs (Generating Evid. Methods to Improv. patient outcomes)*, vol. 5, no. 1, p. 14, Sep. 2017.

De-Identifying Swedish EHR Text Using Public Resources in the General Domain

Taridzo CHOMUTARE^{a,1}, Kassaye Yitbarek YIGZAW^a Andrius BUDRIONIS^a
Alexandra MAKHLYSHEVA^a Fred GODTLIEBSEN^{a,c} and Hercules DALIANIS^{a,b}

^aNorwegian Centre for E-health Research, Tromsø, Norway

^bDepartment of Computer and Systems Sciences, Stockholm University, Sweden

^cFaculty of Science & Technology, UiT - The Arctic University of Norway

Abstract. Sensitive data is normally required to develop rule-based or train machine learning-based models for de-identifying electronic health record (EHR) clinical notes; and this presents important problems for patient privacy. In this study, we add non-sensitive public datasets to EHR training data; (i) scientific medical text and (ii) Wikipedia word vectors. The data, all in Swedish, is used to train a deep learning model using recurrent neural networks. Tests on pseudonymized Swedish EHR clinical notes showed improved precision and recall from 55.62% and 80.02% with the base EHR embedding layer, to 85.01% and 87.15% when Wikipedia word vectors are added. These results suggest that non-sensitive text from the general domain can be used to train robust models for de-identifying Swedish clinical text; and this could be useful in cases where the data is both sensitive and in low-resource languages.

Keywords. EHR, clinical text, de-identification, deep learning, wiki word vectors

1. Introduction

De-identifying health data is an important problem for health data reuse, and the topic has generated significant scholarly interest because of increased use of electronic health records (EHR). Re-use of the data in research could give us unique insights into disease etiology and progression, as well as a greater understanding of patient care processes and pathways. Current de-identification methods rely on sensitive health data for training. This presents a number of data-sensitivity problems, such as when there is need to transfer or adapt the models to new target data. In this study, we investigate the usefulness of non-sensitive training data from the general domain.

Two main approaches have so far been used for de-identification namely, rule-based and machine learning-based methods [1]. Studies show that more successful de-identification systems use a hybrid of both these approaches [2]. On the one hand, rule-based methods can go as far as using name lists from the economy/administration software to match against the clinical text [3]. While this can be an effective solution, it is not robust enough for simple variations or for use outside the specified datasets or organizations, and could entail serious risks to patient privacy. On the other hand, machine

¹Corresponding Author: Taridzo Chomutare; E-mail: firstname.lastname@ehealthresearch.no

learning approaches, while more robust since they learn patterns, instead of matching specific instances, still require a large amount of sensitive data.

Machine learning approaches require a lot of training data or examples to learn from. Creating examples by annotating the data is an expensive proposition because it requires specialist knowledge, and the amounts of data are enormous. Unsupervised methods which can be used to discover discriminating features in new target datasets are emerging. These emerging deep learning architectures do not require any feature engineering to produce state of the art results [4]. So far however, these architectures have only used embeddings from sensitive data or scientific medical publications like PubMed [5]. Out-of-domain sources such as Wikipedia or newer general language models like BERT [6] have not been extensively explored for this task on medical text.

Exploring use of non-sensitive data, the validity of using pseudonymised clinical text for de-identification is studied in [7] where the Stockholm EPR PHI Pseudo Corpus [8] is used and compared with Stockholm EPR PHI Corpus, the non-pseudonymised corpora. It is shown that the results using pseudonymised corpora as training data are slightly decreased, suggesting limited potential.

In another approach, McMurray et al. [3] used both EHR text and text from publicly published medical journals for training purposes. The authors argued medical publications will generally not contain enough protected health information (PHI) information, and this could be a discriminating factor. In contrast, a recent study by Berg et al. [9] found no additional benefits of using out-of-domain training material for de-identification using deep learning approaches.

Whether non-sensitive medical text such as scientific medical publications or even text from the general domain is useful for de-identification, is still a matter without fully resolved clarity. In this study we test both these non-sensitive sources and contribute evidence to help answer the question.

2. Method

Experiments will compare the effect of adding medical scientific text versus text from the general domain to the training set for a de-identification deep learning model. The comparisons are with (i) the base embedding layer from the EHR text, (ii) EHR text plus medical scientific text, and (iii) EHR text plus Wikipedia word vectors. These data sources are detailed in the succeeding subsections.

2.1. Stockholm EPR PHI Psuedo Corpus

Stockholm EPR PHI Pseudo Corpus² is a Swedish EHR corpus, which has been de-identified and pseudonymized [8], and where the tokens are annotated with PHI information. Stockholm EPR PHI Psuedo Corpus is part of the Health Bank [10], the Swedish Health Record Research Bank³. The Health Bank encompasses structured and unstructured patient records data from 512 clinical units from Karolinska University Hospital collected from the years 2007 to 2014 encompassing over 2 million patients. The dataset uses a less fine-grained annotation scheme (IOB), indicating [I=inside token], [B=begin token], and [O = not PHI token].

²Research approved by the Regional Ethical Review Board in Stockholm; permission no. 2014/1607-32.

³Health Bank, <http://www.dsv.su.se/healthbank>

2.2. Scientific medical journal and Swedish Wiki word vectors

Scientific medical text is based on the Läkartidningen corpus (The Swedish scientific medical journal from 1996 to 2005). Läkartidningen has publicly available articles at Språkbanken⁴. Wiki word vectors are pre-trained word vectors created with fastText from Swedish Wikipedia text [11], and are publicly available at fastText⁵. They are designed with no specific downstream task in mind, but what makes them interesting is their use of character-level n-grams, where a single word can be represented by several character n-grams.

2.3. Deep recurrent neural networks

A state of the art deep learning algorithm previously used on health data [5], the Bidirectional Long Short-Term Memory algorithm with conditional random fields (BI-LSTM-CRF), was used in the experiments, as implemented in TensorFlow/Keras⁶. For the scientific medical text, we used another state of the art method, Word2Vec, to create the word embeddings. Wikipedia word vectors are made available to the public pre-trained and ready for downstream tasks. Both sources have 300 dimensional vector representation.

3. Results

The results in Table 1 show a clear improvement in results, from adding Wiki word vectors to the base embedding layer with EHR data only. We also observe that adding scientific medical text improves performance, but falls short of Wiki word vectors.

PHI	EHR			EHR + Scientific medical text			EHR + Wikipedia		
	P %	R %	F1	P %	R %	F1	P %	R %	F1
Age	66.67	40.00	50.00	100.00	80.00	88.89	100.00	80.00	88.89
Date Part	62.87	83.24	71.63	92.09	91.06	91.57	87.76	96.09	91.73
First Name	72.22	87.39	79.09	89.83	66.81	76.63	95.78	95.38	95.58
Full Date	50.00	85.54	63.11	67.23	96.39	79.21	80.41	93.98	86.67
Health Care U.	40.39	77.15	53.02	67.10	77.15	71.78	71.43	82.4	76.52
Last Name	91.61	97.26	94.35	77.01	98.63	86.49	92.95	99.32	96.03
Location	21.15	18.64	19.82	87.50	11.86	20.90	100.00	15.25	26.47
Phone Number	17.39	42.11	24.62	66.67	31.58	42.86	92.86	68.42	78.79
Avg	55.62	80.02	65.62	77.83	77.21	77.52	85.01	87.15	86.07

Table 1. De-identification results based on the three comparisons, P=Precision, R= recall, both percentage

There are a number of reasons that could explain why the Wikipedia text performed better than medical text. First, Wikipedia is a rich source of information which contains both general text and medicine-related text as well. In addition, a number of PHI information such as first and last names, ages, year, and location are present in the text. Also, the scientific medical journal corpus in Swedish (Läkartidningen) produced 118,683 vectors while Wikipedia, on the other hand, produced 1,143,274 vectors.

Further, we observed that the scientific medical text start's out with a relatively high error loss in each epoch, while initial error loss is much lower for Wikipedia. In terms of

⁴Läkartidningen, <https://spraakbanken.gu.se/swe/resurs/lakartidn-vof>
⁵fastText, <https://fasttext.cc>
⁶TensorFlow, <http://www.tensorflow.org>

the improvement in F1 measures (see Figure 1), there was significant performance gain for *Age* and *Phone Number*. For scientific medical text, we noted poorer performance for some PHI information like first names and last names, compared to the EHR baseline.

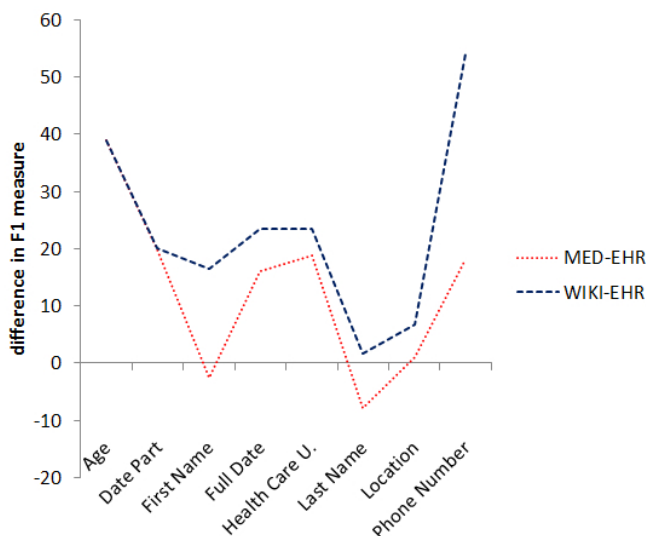


Figure 1. The graph shows the PHI differences in F1 measures between scientific medical text and the EHR baseline (MED-EHR) and between Wikipedia and the EHR baseline (WIKI-EHR) respectively.

4. Discussion

It appears the general consensus in scholarship is that training on general-domain text is not appropriate for tasks on clinical text, since clinical text is so different that it represents a unique linguistic genre. The language in clinical notes is meant for other healthcare professionals. Clinicians and nurses write these notes under time pressure, therefore the text has abbreviations, misspellings, unusual grammatical constructs and other errors and ambiguities.

Our results support a counter-argument that PHI information is distinct from clinical text since PHI information is general, as opposed to clinical procedures, medication or medical concepts that are present in clinical text. Therefore, it could be appropriate to use non-sensitive text in the general domain as training data for detecting PHI information. Also, deep learning architectures have been reported to show good performance under different domains and languages.

The poor results obtained with scientific medical text is consistent with previous assertions made in the literature, that is, scientific text is not likely to contain names and surnames in meaningful contexts [3]. However, the significant improvement in *Age* and *Phone Number* suggest that scientific medical text could still be useful for detecting specific PHI information. Therefore, combining this medical text with other sources could be a viable option.

5. Conclusion

Current results suggest that non-sensitive resources in the general domain can be useful for de-identification tasks on clinical notes. Even though deep learning models are generally thought of as data-hungry, current results raise the prospect of creating robust models; where the primary training data is sensitive and low resourced. In the future, we will test non-sensitive resources and language models to adapt and transfer deep learning models for de-identifying clinical notes between closely similar Nordic languages; such as between Swedish and Norwegian clinical notes.

Acknowledgments

This work is partially supported by the Northern Norway Regional Health Authority, Helse Nord; research grant HNF1395-18.

References

- [1] O. Ferrández, B.R. South, S. Shen, F.J. Friedlin, M.H. Samore and S.M. Meystre, Evaluating current automatic de-identification methods with Veteran's health administration clinical documents, *BMC medical research methodology* **12**(1) (2012), 109.
- [2] A. Dehghan, A. Kovacevic, G. Karystianis, J.A. Keane and G. Nenadic, Combining knowledge-and data-driven methods for de-identification of clinical narratives, *Journal of biomedical informatics* **58** (2015), S53–S59.
- [3] A.J. McMurphy, B. Fitch, G. Savova, I.S. Kohane and B.Y. Reis, Improved de-identification of physician notes through integrative modeling of both public and private medical text, *BMC medical informatics and decision making* **13**(1) (2013), 112.
- [4] F. Dernoncourt, J.Y. Lee, O. Uzunur and P. Szolovits, De-identification of Patient Notes with Recurrent Neural Networks, 2016.
- [5] Z. Liu, B. Tang, X. Wang and Q. Chen, De-identification of clinical notes via recurrent neural network and conditional random field, *Journal of Biomedical Informatics* **75** (2017), S34–S42.
- [6] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *arXiv preprint arXiv:1810.04805* (2018).
- [7] H. Berg, T. Chomutare and H. Dalianis, Building a De-identification System for Real Swedish Clinical Text Using Pseudonymised Clinical Text, in: *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, in conjunction with *Conference on Empirical Methods in Natural Language Processing, (EMNLP) November 2019, Hongkong, ACL., 2019*, pp. 118–125.
- [8] H. Dalianis, Pseudonymisation of Swedish Electronic Patient Records using a rule-based approach, in: *Proceedings of the Workshop on NLP and Pseudonymisation, NoDaLiDa, Turku, Finland September 30, 2019, 2019*.
- [9] H. Berg and H. Dalianis, Augmenting a De-identification System for Swedish Clinical Text Using Open Resources (and Deep learning), in: *Proceedings of the Workshop on NLP and Pseudonymisation, NoDaLiDa, Turku, Finland September 30, 2019*.
- [10] H. Dalianis, A. Henriksson, M. Kvist, S. Velupillai and R. Weegar, HEALTH BANK-A Workbench for Data Science Applications in Healthcare., in: *CAiSE Industry Track*, 2015, pp. 1–18.
- [11] P. Bojanowski, E. Grave, A. Joulin and T. Mikolov, Enriching word vectors with subword information, *Transactions of the Association for Computational Linguistics* **5** (2017), 135–146.

Deep Learning for Accurate Diagnosis of Glaucomatous Optic Neuropathy Using Digital Fundus Image: A Meta-Analysis

Mohaimenul ISLAM^{a,b}, Tahmina Nasrin POLY^{a,b}, Hsuan Chia YANG^{a,b}, Suleman ATIQU^{Ec}; Yu-Chuan Jack LI^{a,b,1}

¹Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan.

²International Center for Health Information and Technology, Taipei Medical University, Taipei, Taiwan.

³College of Public Health and Health Informatics, University of Ha'il, Saudi Arabia

Abstract: We conducted a study to evaluate the algorithms based on deep learning to automatically diagnosis of GON from digital fundus images. A systematic articles search was conducted in PubMed, EMBASE, Google Scholar for the study that investigated the performance of deep learning algorithms for the detection of GON. A total of eight studies were included in this study, of which 5 studies were used to conduct our meta-analysis. The pooled AUROC for detecting GON was 0.98. However, the sensitivity and specificity of deep learning to detect GON were 0.90 (95% CI: 0.90-0.91), and 0.94 (95%CI: 0.93-0.94), respectively.

Keywords: Glaucoma, glaucomatous optic neuropathy, deep learning, artificial intelligence, fundus image.

1. Introduction

Glaucoma has already been emerged as one of primary cause of blindness, affecting approximately 66.8 million people worldwide [1]. It is often considered as a progressiveness neurodegenerative, multifactorial disease, characterized by the continuous loss of ganglion cells (RGCs) and damage of optic nerve atrophy and the visual cortex [2]. Most cases of GON are asymptomatic until to be reached in late stage. Since the diagnosis of GON is very subjective and qualitative, and early detection of GON can help to improve treatment plan, monitoring of the quality of life and to prevent visual impairment. Screening of GON is always achieved by manually assessing optic nerve structure in the digital fundus image. However, this process is labor-intensive, time-consuming and the performance of GON detection solely depend on the skill and experience of ophthalmologists [3]. A fully automated system would enhance the quality of existing care by providing diagnostic support to ophthalmologists and

¹ Yu-Chuan (Jack) Li, MD, PhD Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan. Email: jack@tmu.edu.tw / jaak88@gmail.com

could contribute to the accessibility of patients care by generating large-scale automated screening systems [4].

Deep learning algorithm has already appeared as a cutting-edge solution for successful detection of diabetic retinopathy [5], glioma [6], macular degeneration [7]. An automated system could also play an important role in the diagnosis of GON early that could improve patients care. It would contribute to strengthening the treatment adherence and by providing the importance of GON medications and surgeries. Here, we conducted a systematic review with meta-analysis to evaluate the performance of deep learning for detecting GON using digital fundus images. It can help to establish evidence of the ability of deep learning model for detecting GON.

2. Methods

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) that is based on Cochrane's Handbook for Systematic Reviews [8].

Databases Search A systematic search was conducted in PubMed, EMBASE, Scopus, Google Scholar for relevant studies between 2010 and 2019. Some control keywords: "deep learning", "Convolutional neural network", "CNN", "DL", "glaucoma", "glaucomatous optic neuropathy", "GON" were used as a search term. To ensure comprehensiveness, the reference lists of retrieved articles were checked carefully to get additional relevant studies.

Eligibility Criteria Eligibility was restricted to studies that evaluated the performance of deep learning algorithms to detect glaucoma. Studies were excluded if they were not published in English, and published as review, short report, editorial, letter to editors.

Selection Process Two authors (MMI, TNP) independently screen the titles and abstracts of all retrieved relevant studies. They selected studies based on the pre-specified selection criteria. Any disagreement between them during the screening process was resolved by the main investigator (YCL).

Data Extraction Data collection forms were used to extract the required information from included studies retrieved initially. Two authors first removed duplication by checking author name, publication date, sample size. Afterwards, they collected author name, publication year, total sample size, number of images, methods, external validation information, the performance of deep learning, sensitivity, specificity, true positive, true negative, false positive and false negative.

Assessment of Risk Bias Same two authors independently used QUADAS-2 for assessing the individual quality of each study. The QUADAS-2 tool is recommended for evaluating the risk of bias and applicability of primary diagnostic accuracy studies. QUADAS-2 tool consists of four areas such as patient selection, index test, reference standard and flow and timing.

Statistical Analysis The meta-analysis was performed on qualifying studies that reported the performance of deep learning for detecting GON with sensitivity and specificity, and a total number of images. We then calculated the true positive, false positive, true negative and false negative to generate area under the receiver operating characteristic curve. A random-effect model was used to generate the

pooled sensitivity, specificity, positive- and negative likelihood ratio and diagnostic odds ratio. All the analyses were performed using Meta-Disc (V-1.4) software.

3. Results

The literature searches of the electronic databases yielded 150 articles. After screening of the titles and abstracts, 137 studies were excluded due to lack of inclusion criteria. 13 articles went for full-text review; a total of 8 articles were then selected for our study [3, 9-15]. 5 out of 8 studies were included in the meta-analysis. Publication years ranged from 2018-2019. Three studies were from China, 2 from Japan, 1 from India, 1 from South Korea and 1 from the United States. All the studies were used the CNN model to detect GON from digital fundus images. The QUADAS-2 tool assessment showed that 5 studies had moderate applicability, and the concerns about the applicability of the remaining studies were considered as low.

Table 1: Characteristics of included studies.

Author	Country	Method	Total Image	Data source
Liu 2019	China	CNN	274,413	CGSA*
Asaoka-2019	Japan	CNN	1,364	MRCH
Shibata-2018	Japan	CNN	1,364	MRCH
Li-2018	China	CNN	48,116	LabelMe
Raghavendra- 2018	India	CNN	1,426	KMC
Ahn-2018	S Korea	CNN	1,542	KEH
Christopher- 2018	USA	CNN	14,822	ADGES & UCSD
Chai-2018	China	CNN	2,554	NR

Note= CGSA: Chinese Glaucoma Study Alliance *other dataset also used. MRCH: Matsue Red Cross Hospital, KMC= Kasturba Medical College, Kim's Eye Hospital, The African Descent and Glaucoma Evaluation Study and the University of California, San Diego (UCSD), NR= Not reported.

Five studies evaluated the performance of deep learning model for accurate detection of GON. The deep learning model showed the higher performance to accurately detect GON. The overall pooled AUROC for deep learning algorithms to detect GON was 0.98 (**Figure 1**). The sensitivity and specificity were 0.90 (95% CI: 0.90-0.91), and 0.94 (95%CI: 0.93-0.94), respectively.

Furthermore, the pooled positive likelihood ratio and negative likelihood ratio were 16.77 (95%CI: 6.42-43.79), and 0.06 (95%CI: 0.02-0.15). The heterogeneity among the studies was $I^2=99.8\%$ and $I^2= 99.4\%$, respectively. Moreover, the diagnostic odds ratio was 292.53 (50.13-1706.94).

4. Discussion

In this study, we evaluated the performance of deep learning algorithms for detecting GON from the digital fundus images. Our findings suggest that deep learning model has immense potential to detect GON with high sensitivity and specificity in a diverse population. The diagnosis accuracy, sensitivity and specificity of traditional computer

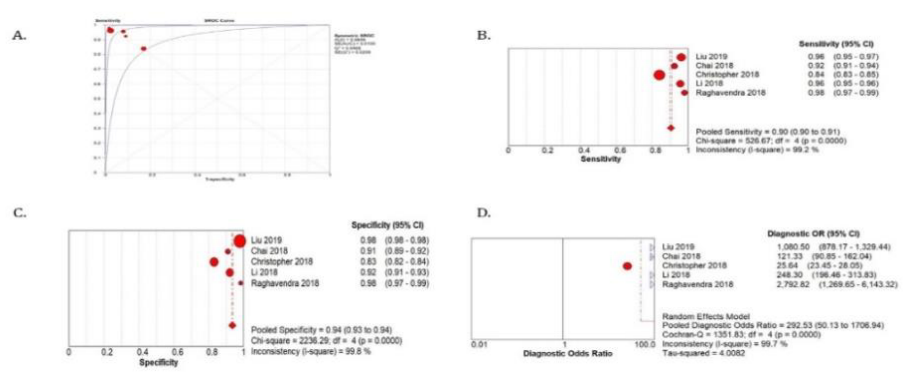


Figure 1: Classification performance of deep learning model. a) Receiver operating characteristic curve b) sensitivity c) specificity d) diagnostic odds ratio

vision methods are close to 60% [16], whereas the pooled AUROC, sensitivity, and specificity of deep learning model are 0.98, 0.90, and 0.94, respectively. The performance of the deep learning model can help to make an automatic screening tool to detect GON early for better patient's care. Improved sensitivity and specificity of tests increase the feasibility of GON screening programs by reducing the burden of false-positive [12]. Early diagnosis of GON using automated tools could have a crucial role in supporting ophthalmologist, as patients with glaucoma are unaware of their condition until central visual acuity is damaged. An automated technique would help to improve performance, reduce the healthcare cost, and give the same performance in different digital images.

There are several advantages of using deep learning for the detection of GON from digital fundus images. First, no need to do conventional steps such as feature extraction, dimensionality reduction and feature ranking [17]. Second, it does not require data preprocessing which usually influence in traditional machine learning techniques. However, there are several challenges to use deep learning techniques in GON diagnosis. First, it takes a higher number of GON images to train the model for getting the optimum performance. Second, generalizability problem, it means the model would perform the same in different databases.

5. Conclusion

Findings of our current study showed that deep learning model has immense potential to detect GON (high sensitivity and specificity). Implementation of deep learning-based automated tools could be cost-effective and improve the accessibility of GON screening in a remote setting. Having ensured the generalization of automated diagnosis of GON, a dataset with different ethnicities could be tested.

References

[1] Shahsuvaryan ML. Glaucomatous optic neuropathy management: the role of neuroprotective agents. Medical hypothesis, discovery and innovation in ophthalmology 2013;2(2):41.

- [2] Kaushik S, Pandav S, Ram J. Neuroprotection in glaucoma. *Journal of postgraduate medicine* 2003;49(1):90.
- [3] Liu H, Li L, Wormstone IM, Qiao C, Zhang C, Liu P, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. *JAMA ophthalmology* 2019.
- [4] Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA ophthalmology* 2018;136(7):803-10.
- [5] Poly TN, Islam MM, Yang HC, Nguyen PA, Wu CC, Li Y. Artificial Intelligence in Diabetic Retinopathy: Insights from a Meta-Analysis of Deep Learning. *17th World Congress on Medical and Health Informatics, MEDINFO 2019*. IOS Press; 2019:1556-7.
- [6] Ertosun MG, Rubin DL. Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks. *AMIA Annual Symposium Proceedings*. 2015. American Medical Informatics Association; 2015:1899.
- [7] Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA ophthalmology* 2017;135(11):1170-6.
- [8] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 2009;151(4):264-9.
- [9] Ahn JM, Kim S, Ahn K-S, Cho S-H, Lee KB, Kim US. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLoS One* 2018;13(11):e0207982.
- [10] Asaoka R, Tanito M, Shibata N, Mitsuhashi K, Nakahara K, Fujino Y, et al. Validation of a deep learning model to screen for glaucoma using images from different fundus cameras and data augmentation. *Ophthalmology Glaucoma* 2019.
- [11] Chai Y, Liu H, Xu J. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. *Knowledge-Based Systems* 2018;161:147-56.
- [12] Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, et al. Performance of Deep Learning Architectures and Transfer Learning for Detecting Glaucomatous Optic Neuropathy in Fundus Photographs. *Scientific reports* 2018;8(1):16685.
- [13] Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* 2018;125(8):1199-206.
- [14] Raghavendra U, Fujita H, Bhandary SV, Gudigar A, Tan JH, Acharya UR. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Information Sciences* 2018;441:41-9.
- [15] Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Scientific reports* 2018;8(1):14665.
- [16] Mookiah MRK, Acharya UR, Lim CM, Petznick A, Suri JS. Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. *Knowledge-Based Systems* 2012;33:73-82.
- [17] Islam MM, Poly TN, Li Y-CJ. Retinal Vessels Detection Using Convolutional Neural Networks in Fundus Images. *bioRxiv* 2019:737668.

Design for a Modular Clinical Trial Recruitment Support System Based on FHIR and OMOP

Ines REINECKE^{a,1}, Christian GULDEN^b, Michéle KÜMMEL^a,
Azadeh NASSIRIAN^a, Romina BLASINI^c, and Martin SEDLMAYR^a

^a*Carl Gustav Carus Faculty of Medicine, Center for Medical Informatics, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany*

^b*Medical Informatics, Univ. of Erlangen-Nürnberg, Erlangen, Germany*

^c*Medical Informatics, Univ. of Giessen, Giessen, Germany*

Abstract. The MIRACUM consortium is developing a Clinical Trials Recruitment Support System to support the data-driven recruitment of patients for clinical trials. The design of the prototype includes both open source solutions (OMOP CDM, Atlas) and open standards for interoperability (FHIR). The aim of the prototype is to create a patient screening list of potential participants for a clinical study. The paper shows the modular structure and functionality of the prototype building the foundation for the practical implementation of the CTRSS and, at the same time, demonstrating the use of open source solutions and standards for the development of clinical support systems.

Keywords. clinical trial recruitment support system, FHIR, OMOP, OHDSI, secondary use of electronic health records, automation, open standards, decision support systems

1. Introduction

Clinical trials are fundamental in clinical research and in evidence based medicine since they are needed to evaluate new treatment methods, the effectivity of drugs, or new diagnostic tools [1]. However, many studies fail during the recruitment phase for the following two reasons: either the minimum necessary number of patients is not reached or the recruitment process takes too much time and effort to find enough eligible patients to participate [2] [3]. One option to overcome those issues is the specific reuse of already observed data during inpatient treatment. Since the availability of digital health care data is growing rapidly it offers high potential for reuse in clinical research [4]. The idea of automated screening electronic health care data is not new. During the last ten years the number of publications on clinical trial recruitment support systems (CTRSS)

¹Corresponding Author: Ines Reinecke, Teamleader Research Infrastructure, Research Associate, Fetscherstrasse 74, 01307 Dresden, Germany; E-mail: ines.reinecke@uniklinikum-dresden.de

is continuously increasing. This trend demonstrates the importance and need of such systems to allow faster and more efficient patient screening [5]. Köpcke et al. [5] pointed out the lack of standards and missing interoperability of CTRSS as well as a missing formalisation of eligibility criteria for the definition of a potential trial cohort. Thus a major goal of the MIRACUM consortium [6] is to provide a modular CTRSS based on open standards. This approach shall ensure interoperability and flexibility to connect to other systems such as study registries and hospital information systems. As one of four consortia MIRACUM is part of the German Medical Informatics Initiative (MII). The MII brings together 30 university hospitals and a number of industry partners to establish data integration centres. The aim is to provide clinical health care data from various source systems in one place to ensure to "prove their benefit for researchers, doctors and patients" [7]. To demonstrate the benefits of integrated data one use case the MIRACUM consortium is responsible for is "Alerting in care - IT support for patient recruitment". This paper illustrates how to build a CTRSS concept that consists of separate modules, can be transferred to other locations, make use of existing open source tools and benefit from open standards.

2. Methods

To address the topics modularity, transferability, open source tools and open standards, a heterogeneous group of computer scientists, medical information scientists, data scientists and physicians joined forces to develop the CTRSS prototype. The team evaluated how the communication between loose coupled modules of the prototype can be done based on Representational State Transfer (REST) over HTTP protocol and what is needed to represent data using the Fast Healthcare Interoperability Resources (FHIR) developed by Health Level Seven Organization (HL7) standard for data exchange. A common data model was needed to store data that gets searched for potential trial participants by the CTRSS prototype. The *Observational Health Data Sciences and Informatics* (OHDSI) organization consists of a rapidly growing worldwide community of universities to foster worldwide medical scientific community and to improve scientific research [8]. Thus they developed a common data model (CDM) *Observational Medical Outcomes Partnership* (OMOP) including a set of standardized vocabularies (e.g. SNOMED, LOINC, rxNorm, ICD10) and their translations between each other to allow researchers to speak a common language when exchanging data. To validate OMOP CDM is a proper patient data storage needed by a CTRSS we installed it on a PostgreSQL database and added standardised vocabularies using the OHDSI Athena web application. Athena is a service that provides regular updates of the vocabularies to ensure utilization of the most recent vocabulary versions. In Germany a modified version of the ICD10 world health organization (WHO) vocabulary is in place to serve local billing purposes. Thus the ICD10GM vocabulary has been added as an additional vocabulary to the OMOP database based on the work done by Maier et al. [9]. In a next step the patient data based on the MII core data set were imported into OMOP. All data included in the MII core data have been mapped into the OMOP CDM successfully. Afterwards the Atlas web application was used to define cohorts based on the eligibility criteria of a clinical trial using a web interface for several clinical trials. The exchange of cohort definitions between different locations was done with the import and export capability of the Atlas

web application. All information of defined cohorts was stored in the OMOP database. Furthermore OHDSI provides a Web Application Programming Interface (API) based on REST for the OMOP CDM. The REST Web API was used to realize the access to already defined cohorts. This included the execution of cohort definitions against the OMOP database, status checks of a cohort execution and the call of a result list of patient ids for a certain cohort definition. To implement the notification of research teams about any changes on the patient screening lists, we evaluated the functionality offered by the FHIR HL7 standard. FHIR provides subscription resources that allows us to register a certain resource to act on any event such as change, create or delete. The subscription resource was configured to use a certain communication channel such as e-mail, REST post request and others.

3. Results

The CTRSS prototype consists of separate modules based on an architecture proposed by Trinczek et. al. [10]. Each module targets a small set of similar features and is implemented as a REST based micro-service with well-defined service boundaries. All module interfaces have a properly documented API. As shown in Figure 1 the prototype CTRSS is composed of the following five modules:

- patient data module
- Atlas web application
- query module
- notification module
- screening list module

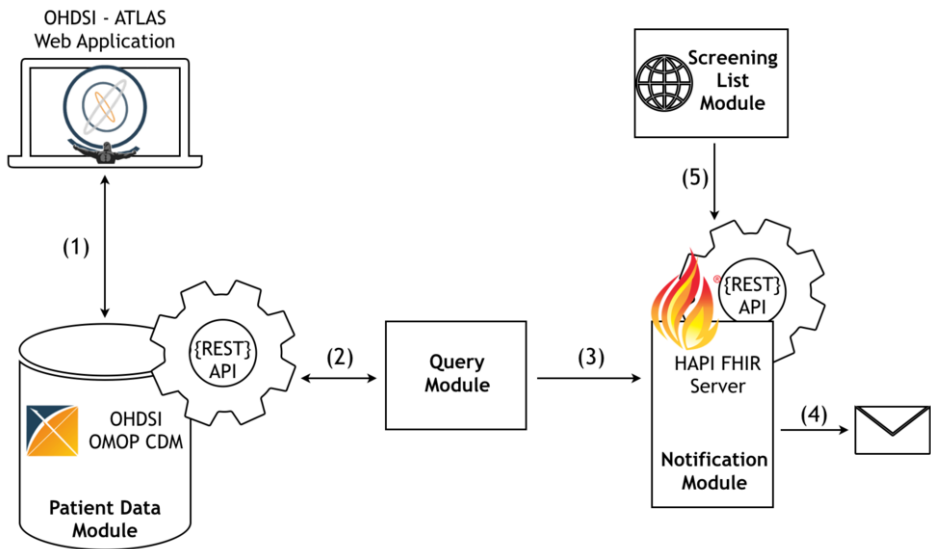


Figure 1. Architecture of the CTRSS

The process of creating a patient screening list consists of the following five main steps:

1. cohort definition using the ATLAS web application
2. generation of a cohort and retrieval of its contents from the OMOP database done by the query module
3. generation or update of the screening list with retrieved patients as a FHIR list-resource done by the query module
4. notification based on any event on a FHIR list-resource done by the notification module
5. screening list module receives new or changed screening lists

The patient data module is based on the OMOP CDM (version 5) and it stores all observational data of a patient. In the first step the Atlas web application is being used for cohort definition. On successful cohort definition an unique identifier of the defined cohort is returned. All cohort definitions built with Atlas can be exported to other locations. The query module is a custom implementation using REST calls to send and retrieve data. In step two the query module generates the patient list using the cohort identifier that is handed over after cohort definition with Atlas. Based on a defined scheduling or on notification when patient data changes, the query module runs a REST POST query against the OMOP Web API to generate the list of potential trial participants in the database. The query module uses another REST call to check the current status of a cohort generation. After successful cohort creation the query module retrieves the patient list. In step three the list of patients for a particular trial is sent using a FHIR list-resource that contains all patient ids. There has been a subscription resource created for the list-resource that sends e-mail messages to recipients when a list of patient ids gets created or changed on the FHIR server. The patient screening list module is a custom web application that reads the list-resources data from the FHIR server. It displays all potential patients for a trial of interest based on the cohort id. The screening list module is implemented as a web application based on HTML 5 and the java script framework Vue.js.

4. Discussion

The CTRSS prototype is based on separate modules that consist of a strong encapsulation to hide the details of the implementation that ensure a loose coupling between each other and thus addresses the lack of missing interoperability [11]. Our approach provides great advantages in terms of technology heterogeneity, resilience, scalability, deployment efficiency, composability, and optimization for replacement as stated by Sam Newman [12]. Hence compared to monolithic software, the modular CTRSS prototype introduced by us allows to accelerate development cycles, reduce deployment efforts, and provide improved scalability. Development teams can exchange single modules without affecting others as long as the interfaces are the same and thus flexibility of the whole system increases. Additionally the scaling of each module can be done based on performance requirements individually instead of scaling the whole prototype. As stated by Köpcke et. al. [5] "many CTRSS are designed to fit the existing infrastructure of a clinical care provider or the particularities of a trial" and therefore most existing CTRSS are limited to local environments and conditions that cannot be transferred to other locations nor adapted to trials with different characteristics. Our CTRSS prototype was built to over-

come those limitations as it is a transferable solution that can be used across different locations. Additionally the usage of FHIR for the exchange of electronic healthcare data using web technologies enables the prototype to be extended and connected to many other health care systems. In the future the prototype can be extended to send back notification on new potential trial participants to existing health information systems (HIS). With the OMOP CDM the used patient data is transformed into a standardized format that allows scientists to search for potential participants at multiple locations national wide and with the option for extension to an international level in future projects. The usage of OHDSI Atlas allows scientists to share defined cohorts with others. This extends the range of coverage when searching for potential participants and thus increases the chance to meet set expectations in terms of number of participants. Our prototype is currently limited in terms of usage and result reporting. Therefore no results on the success of patient recruitment can be provided. The exchange of data between different locations has to be done in a next step. The test and deployment of the prototype within the MIRACUM consortium including reporting of the results and outcomes needs to be addressed in follow-up publication as suggested by Köpcke et. al. [5].

Acknowledgement

This work is part of the project MIRACUM, funded by the German Ministry of Education and Research (FKZ 01ZZ1801A/L).

References

- [1] Backmann M. What's in a gold standard? In defence of randomised controlled trials. *Medicine, Health Care and Philosophy*. 2017 apr;20(4):513–523.
- [2] Ohmann C, Kuchinke W. Meeting the Challenges of Patient Recruitment. *International Journal of Pharmaceutical Medicine*. 2007;21(4):263–270.
- [3] Sully BGO, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. *Trials*. 2013;14(1):166.
- [4] Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. *Yearbook of Medical Informatics*. 2017;26(01):38–52.
- [5] Köpcke F, Prokosch HU. Employing Computers for the Recruitment into Clinical Trials: A Comprehensive Systematic Review. *Journal of Medical Internet Research*. 2014 jul;16(7):e161.
- [6] Prokosch HU, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. *Methods of Information in Medicine*. 2018 jul;57(S 01):e82–e91.
- [7] Gehring S, Eulenfeld R. German Medical Informatics Initiative: Unlocking Data for Research and Health Care. *Methods of Information in Medicine*. 2018 jul;57(S 01):e46–e49.
- [8] Sciences OHD, Informatics. OHDSI The Book of. Independently published; 2019.
- [9] Maier C, Lang L, Storf H, Vormstein P, Bieber R, Bernarding J, et al. Towards Implementation of OMOP in a German University Hospital Consortium. *Applied Clinical Informatics*. 2018 jan;09(01):054–061.
- [10] Trinczek B, Köpcke F, Leusch T, Majeed RW, Schreibeis B, Wenk J, et al. Design and multicentric Implementation of a generic Software Architecture for Patient Recruitment Systems re-using existing HIS tools and Routine Patient Data. *Applied Clinical Informatics*. 2014;05(01):264–283.
- [11] Krafzig D. Enterprise SOA: Service-Oriented Architecture Best Practices. Prentice Hall; 2004.
- [12] Newman S. Building Microservices: Designing Fine-Grained Systems. O'Reilly Media; 2015.

Detection of Muscle Weakness in Medical Texts Using Natural Language Processing

Gleb DANILOV^{a,1}, Michael SHIFRIN^a, Yuliya STRUNINA^a, Konstantin KOTIK^a,
Tatyana TSUKANOVA^a, Tatiana PRONKINA^a,
Timur ISHANKULOV^a, Elizaveta MAKASHOVA^b, Alexandra KOSYRKOVA^a and
Alexander POTAPOV^a

^a *Laboratory of Biomedical Informatics and Artificial Intelligence, National Medical Research Center for Neurosurgery named after N.N. Burdenko, Moscow, Russian Federation*

^b *I.M. Sechenov First Moscow State Medical University, Moscow, Russian Federation*

Abstract. Identifying adverse events in clinical documents is demanded in retrospective clinical research and prospective monitoring of treatment safety and cost-effectiveness. We proposed and evaluated a few methods of semi-automated muscle weakness detection in preoperative clinical notes for a larger project on predicting paresis by images. The combination of semi-expert and machine learning methods demonstrated maximized sensitivity = 0.860 and specificity = 0.919, and largest AUC = 0.943 with a 95% CI [0.874; 0.991], outperforming each method used individually. Our approaches are expected to be effective for autoshaping a well-verified training dataset for supervised machine learning.

Keywords. Electronic Health Records, Neurosurgery, Natural Language Processing, Adverse Events, Logistic Regression, ROC-analysis

1. Introduction

Identifying adverse events in clinical documents is highly demanded in retrospective clinical research and monitoring of treatment safety and cost-effectiveness [1]. However, these entities are described using free text in medical records. To extract information on complications following neurosurgery, the examination of narrative medical texts is likely needed. These activities are time-consuming for experts and almost impractical when dealing with big data collected through decades. Despite the effective machine learning implications in natural language processing, the absence of formalized ground truth variables in unstructured data makes the use of supervised techniques challenging. Thus, the text annotation would benefit from effective automation. Recently, our group proposed an information extraction algorithm (IEA) for detecting adverse events in neurosurgery using documents written in a natural rich-in-morphology language [2]. This study aimed to test our IEA in the identification of preoperative extremity muscle weakness in patients with glial tumors and to compare its feasibility with supervised machine learning.

¹ Corresponding author, E-mail: glebda@yandex.ru.

2. Methods

The task of any extremity weakness (paresis) detection in medical texts originated from the project on muscle strength prediction by deep learning on medical images (supported by Russian Foundation for Basic Research (RFBR) grant 19-29-01154). The technical problem we tried to solve was a classification of patients as having paresis or not. The relevant data were obtained for a cohort of 429 patients with accessible magnetic resonance images (average age 39.8 ± 17.9 y.o., 194 (45.2%) males) who underwent glial tumor resection at N.N. Burdenko National Medical Research Center of Neurosurgery in a period between 2009 and 2018. In this sample, the narrative description of the preoperative neurological status in the Russian language was supplemented with the quantitative (0 to 5 scores) estimation of muscle strength for each extremity, which allowed us to create a ground truth binary variable (0 – no paresis, 1 - any paresis) [3]. That dataset was used in testing of our algorithm's performance and comparing it to supervised machine learning. All the data extracted from electronic health records (EHR) were processed and analyzed within the R programming environment (version 3.5.0) in RStudio IDE for MacOS (version 1.1.453) using *rsample*, *drlib*, *broom*, *yardstick*, *glmnet*, *doMC*, *dplyr*, *tidyr*, *stringr*, *tidytext*, *tidyverse*, *quanteda*, *qdapRegex*, *tm*, *scales*, and *cutpointr* packages.

2.1. Event detection in EHR

Source text documents related to neurological examination upon admission were selected to build the initial corpus of 429 medical cases, which was preprocessed as follows:

- words transformed to lower case;
- all the characters and symbols except for letters and single spaces removed;
- texts tokenized with a space separator;
- stop-words and meaningless words (single letters, artifacts, etc.) removed;
- spelling corrected and tokens lemmatized.

The unique lemmas were screened by an expert to select those likely related to or certainly used in the paresis description. The selected lemmas shaped the “paresis lexicon” list. The context of each selected term was defined from *n*-grams containing it. Considering the importance of prepositions in the Russian language, we have seen 5-grams sufficiently interpretable. Those 5-grams were rated by 0.5 (if indicating any paresis with uncertainty), 1 (if indicating any paresis obviously), or 0 (if no paresis was mentioned) by the expert. We tested two modes of using the scores to decide whether muscle weakness is present in the source documents containing the rated 5-grams. The first option considered paresis in only those medical records in which at least one 5-gram received 1 point (IEA1, “utmost confident algorithm”). In the second mode, the presence of paresis was decided considering the sum of scores for all rated 5-grams found in each case history based on an optimal cutpoint (IEA2, “highly suspicious algorithm”).

2.2. Logistic regression with LASSO regularization

The subset of 429 clinical records containing a narrative and digital preoperative assessment of muscle strength was used to build a supervised logistic regression model

(LR1) on a “bag of words” text representations (sparse vectors). The target binary variable indicated the presence/absence of paresis. The model was evaluated with 5-fold validation (within training data consisted of 75% of the initial sample) and LASSO regularization, 300 times repeatedly using resampling. We rebuilt the LR1 model with the same technique adding more textual data referring to preoperative neurological status, however, written postoperatively (LR2). Finally, IEA1 was used in combination with LR2 to find whether it improves the prediction (IEA1+LR2). In the latter approach, all the cases of paresis confidently identified by IEA1 were automatically accepted, and LR2 predicted the rest.

2.3. Performance evaluation

The quality of the IEA, the logistic regression models, and their combination was evaluated using ROC-analysis. The sensitivity (SENS, also referred to as recall), specificity (SPEC), accuracy (ACC), positive predictive value (PPV, also referred to as precision), negative predicted value (NPV), F-measure (F), cutoff value (COV) and area under ROC-curve (AUC) were computed for all methods on test sets in each iteration of resampling. The bounds of 95% AUC confidence interval were calculated at each iteration using bootstrapping techniques. The cutoff point was chosen on the training sets to maximize the sum of sensitivity and specificity for every method. All the metrics were averaged across 300 iterations.

3. Results

3.1. Evaluation of the paresis detection algorithms

We selected 122 terms to shape “paresis lexicon” out of unique 854 lemmas for 1199 original tokens from 429 cases. These 122 concepts were represented in 491 5-grams screened by the expert. The paresis was suspected with uncertainty by 72 and more confidently by 92 n-grams. The IEA1 showed the best accuracy, positive predictive value, and specificity operating with an only binary (0 or 1) output (Table 1). The output of IEA2 ranged from 0 to 8.5. However, the cutoff point maximizing the sum of sensitivity and specificity treated all cases of 0.5 scores and above as paresis (Table 1). AUC for both cases was estimated above 0.8 on average, with IEA2 outperforming IEA1 (Table 1, Figure 1, A).

Table 1. The quality metrics of information extraction algorithms, logistic regression models and combined approach.

Method	COV*	SENS	SPEC	ACC	PPV	NPV	F	ROC AUC [95% CI]
IEA1	1	0.633	1	0.907	1	0.889	0.775	0.817 [0.771; 0.863]
IEA2	0.5	0.789	0.897	0.869	0.723	0.926	0.754	0.875 [0.833; 0.914]
LR1	0.267	0.668	0.906	0.846	0.723	0.892	0.673	0.860 [0.757; 0.945]
LR2	0.272	0.762	0.925	0.877	0.815	0.904	0.777	0.911 [0.831; 0.972]
LR2 + IEA1	0.268	0.860	0.919	0.902	0.821	0.942	0.837	0.943 [0.874; 0.991]

* for all algorithms, ‘paresis’ cases were considered when output \geq cutoff point.

3.2. Evaluation of logistic regression and combined approach

The LR1 and LR2 algorithms showed better performance than semi-automated expert dependent IEA1 and IEA2, with LR2 superior to LR1 (Table 1, Figure 1, B and C respectively). Regularization led to the selection of words most significant for paresis detection. However, the combination of IEA1 and LR2 simultaneously resulted in the largest AUC in our tests (Table 1, Figure 1, D).

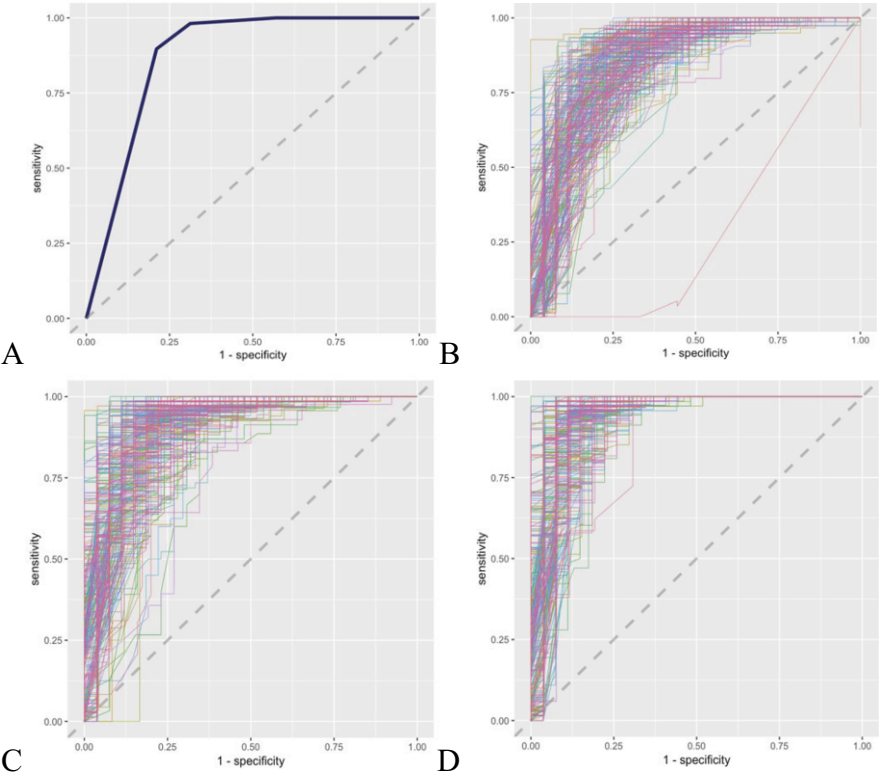


Figure 1. ROC curves superimposed after 300 times of resampling. A – model IEA2, B – model LR1, C – model LR2, D – model IEA1+LR2. The improvement of performance is observed from A to D.

4. Discussion

Natural language processing with machine learning machine learning has occasionally been applied in neurosurgical adverse events detection. J. Gaebel et al. (2015) suggested using semantic preprocessing for adverse events extraction in neurosurgery [4]. B. Campillo-Gimenez et al. (2013) identified surgical site infections with the NOMINDEX tool and text similarity in vector space [5]. N. Tvardik et al. (2018) detected healthcare-associated infections in clinical notes [6]. However, the idea of adverse events extraction from texts is challenged in other medical domains (e.g., infections or drug trials) and with other technologies (e.g., word embeddings, neural networks, etc.) [1], [7].

In our study, the IEA1 method appeared to be robust in the detection of paresis

with 100% specificity and positive predictive value (which still needs verification on a larger dataset). Despite the moderate sensitivity, this semi-expert method might be required in retrospective research when a sufficient number of ‘paresis’ cases should be obtained for clinical study, and no quantified muscle strength assessment is provided for supervised machine learning. However, collecting more data available for IEA1 and a combination of it with logistic regression provides a more flexible instrument to enroll true positive cases and true negative cases in the main and control study groups with stronger confidence at a large number. The other implication of these methods is in prospective adverse events monitoring. IEA1+LR2 favoring maximal sensitivity might be preferable. In general, we consider this methodology feasible for other adverse events detection. IEA1 and IEA2 may be especially efficient in detecting rare cases.

The most important limitation of this work is a relatively small dataset. All methods described above will be further tested on a significantly larger data volume. It would also be valuable to estimate all the models on the dataset with no quantitative muscle strength assessment. The more resource-intensive expert verification will be required in future studies.

5. Conclusion

The method for paresis detection in unstructured medical records showed its efficiency, however testing on a larger dataset is required. The event detection algorithms based on the combination of expertise and machine learning outperformed each approach used independently.

This project was supported by the RFBR grant 18-29-22085. The extraction of data for glial tumors was also supported by the RFBR grant 18-29-01052.

References

- [1] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi,
- [2] S. Sohn, H. Liu, Clinical information extraction applications: A literature review, *Journal of Biomedical Informatics*, **77** (2018), 34–49.
- [3] G. Danilov, M. Shifrin, U. Strunina, T. Pronkina, A. Potapov, An Information Extraction Algorithm for Detecting Adverse Events in Neurosurgery Using Documents Written in a Natural Rich-in-Morphology Language, *Stud. Health Technol. Inform.*, **262** (2019), 194–197.
- [4] A. Compston, Aids to the Investigation of Peripheral Nerve Injuries. Medical Research Council: Nerve Injuries Research Committee. His Majesty's Stationery Office: 1942; pp. 48 (iii) and 74 figures and 7 diagrams; with Aids to the Examination of the Peripheral Nervous System. By Michael O'Brien for the Guarantors of Brain. Saunders Elsevier, 2010 ; pp. [8] 64 and 94 Figures, *Brain*, **133** (2010), 2838–2844.
- [5] J. Gaebel, T. Kolter, F. Arlt, K. Denecke, Extraction Of Adverse Events From Clinical Documents To Support Decision Making Using Semantic Preprocessing, *Stud. Health Technol. Inform.*, **216** (2015), 1030.
- [6] B. Campillo-Gimenez, N. Garcelon, P. Jarno, J. M. Chaplain, M. Cuggia, Full-text automated detection of surgical site infections secondary to neurosurgery in Rennes, France, *Stud. Health Technol. Inform.*, **192** (2013), 572–575.
- [7] N. Tvardik, I. Kergourlay, A. Bittar, F. Segond, S. Darmoni, M.-H. Metzger, Accuracy of using natural language processing methods for identifying healthcare-associated infections, *Int. J. Med. Inform.*, **117** (2018), 96–102.
- [8] F. Shen, et al., Detection of Surgical Site Infection Utilizing Automated Feature Generation in Clinical Notes, *J. Healthc. Informatics Res.*, **3** (2019), 267–282.

Digitalisation of the Brief Visuospatial Memory Test-Revised and Evaluation with a Machine Learning Algorithm

Martin Eduard BIRCHMEIER ^{*, a, 1}, Tobias STUDER ^{*, a}, Prof. Dr. med. Andreas LUTTEROTTI ^b, Prof. Dr. Dipl.-Psych. Iris-Katharina PENNER ^c, Prof. Serge BIGNENS ^a

** Those authors contributed equally to the work*

^a Bern University of Applied Sciences, Biel, Switzerland

^b Department of Neurology, University Hospital Zurich, Zurich, Switzerland

^c COGITO Center, Neurocognition and Neuropsychology, Düsseldorf, Germany

Abstract. The disease multiple sclerosis (MS) is characterized by various neurological symptoms. This paper deals with a novel tool to assess cognitive dysfunction. The Brief Visuospatial Memory Test-Revised (BVM-T-R) is a recognized method to measure optical recognition deficits and their progression. Typically, the test is carried out on paper. We present a way to make this process more efficient, without losing quality by having the patients using a tablet App and having the drawings rated with the use of a machine learning (ML) algorithm. A dataset of 1'525 drawings were digitalized and then randomly split in a training dataset and in a test dataset. In addition to the training dataset the already trained drawings from a preliminary paper were added to the training dataset. The ratings done by two neuropsychologists matched for 81% of the test dataset. The ratings done automatically with the ML algorithm matched 72% with the ones of the first neuropsychologist and 79% of the ones of the second neuropsychologist. For a semi-automated rating we defined a threshold value for the reliability of the rating of 78.8%, under which the drawing is routed for manual rating. With this threshold value the ML algorithm matched 80.3% and 86.6% of the ratings of the first and second neuropsychologists. The neuropsychologists have in that case to manually check 17.4% of the drawings. With our results it is possible to execute the BVM-T-R Test in a digital way. We found out, that our ML algorithms have with the semi-automated method the similar matching as the two professional raters.

Keywords. BICAMS; BVM-T-R; Convolutional neural network; Machine Learning, Multiple Sclerosis, digitalize

1. Introduction

The disease multiple sclerosis (MS) is characterized by various neurological symptoms, including motor weakness and spasticity, coordination problems, fatigue, sensory, bladder and cognitive dysfunction [1]. This paper deals with a novel tool to assess cognitive dysfunction. The Brief Visuospatial Memory Test-Revised (BVM-T-R) [2], which is part of the Brief International Cognitive Assessment for MS Test Battery

¹ Bern University of Applied Sciences Engineering and Information Technology, Höhweg 80, CH-2502 Biel / Bienne, E-mail: m.birchmeier@hotmail.com

(BICAMS) is a validated instrument to assess cognition in MS patients [3]. Currently, the BVMT-R is carried out on paper in the presence of a neuropsychologist. Briefly, the patient is presented a page with six geometric figures for ten seconds to memorize them. Then he tries to redraw them from memory considering the correctness and position of each figure. These drawings are then evaluated by a neuropsychologist. A figure, which is both correct in details and position on the paper receives a rating of 2 points. If the drawing is not correct but similar to the original or correct but in the wrong position, the rating is 1. If the drawing is wrong the rating is 0 points. This is a time-consuming procedure for both players.







In this paper, we present a way to make this process more efficient, without losing quality by having the patients using a tablet App to look at, memorize and draw the figures and having the drawings rated automatically or semi-automatically with the use of a machine learning (ML) algorithm. We have therefore defined the following research question: “Can the drawings of the BVMT-R be rated by a Convolutional Neural Network (CNN) with the same accuracy as a neuropsychologist?”.

2. Method

First of all, a mobile app was developed to digitalize the 294 (page of six drawing of geometric figures) paper-based patient data, which were provided by the COGITO GmbH in Düsseldorf and evaluated by their neuropsychologist (N1). Those patient data contain only the form A (see Table 1 below) of the BVMT-R. After eliminating 239 empty drawings, from the total of 294 patient’s data at 6 drawings per page, a total of 1’525 drawings remained and were included in this patient dataset.

This dataset was then randomly split in 1’220 drawings (=80%) for the training dataset and 305 (=20%) for the test dataset. In addition to the training dataset the already trained drawings from the preliminary paper [4] were added, which resulted in a total training dataset of 1’790 drawings (see table 1).

Table 1. Number of training data (m) and test data (n) per figure.

Number	Figure	Data
1		m = 352, n = 57
2		m = 309, n = 52
3		m = 320, n = 54
4		m = 267, n = 47
5		m = 281, n = 48
6		m = 261, n = 47
Total		m = 1’790, n = 305

For the ML algorithm a CNN was chosen, because this algorithm has been developed for visual object classifications [5]. The CNN gives as an output a rating of the image with a probability value of the reliability of the rating [6].

Furthermore, the test dataset was reevaluated by a second neuropsychologist (N2) from the University Hospital Zurich, to compare the evaluation between two neuropsychologists from different centers. The evaluation of the neuropsychologist from the University Hospital Zurich is also used as external validation unit.

Since the manual rating procedure of a single neuropsychologist is not representing a clear gold standard for the ML algorithm, its rating results cannot rely on the ratings of a single neuropsychologist. Therefore, in order to answer the research question, we use the result of the comparisons between the ML algorithm and the two individual neuropsychologists and consider these results with the comparison between both neuropsychologists among themselves.

Various statistical methods were used to evaluate the ML algorithm: Sensitivity, specificity, the positive and negative prediction value [7], the significance test [8] and the Cohen's kappa test to evaluate the interrater reliability of assessments between the neuropsychologists themselves and the ML algorithm [9].

In a second step, a semi-automated rating methodology was used, keeping the rating of the ML algorithm when the rating reliability was above a defined threshold value and otherwise redirecting the drawing to the neuropsychologist for manual rating.

3. Results

The two neuropsychologists (N1 and N2) rated the test data and they were compared with each other. Their rating was identical for 81% of the drawings. Cohen's Kappa (K) was 0.62.

We can see that our neuropsychologists have an equal match with a Kappa of 0.62 like the scientists of Brazilian study [10].

The agreement between N1 and ML is by 72% ($K = 0.45$) and between N2 and ML 79% ($K = 0.56$).

With the second method (semi-automated rating) in order to set the threshold value, a ROC curve was generated per label and comparison from the sensitivity and the specificity.

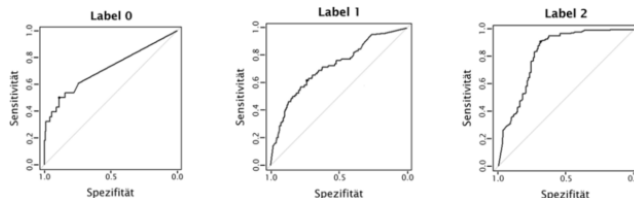


Figure 1. ROC curve per label from the comparison between N1 and ML.

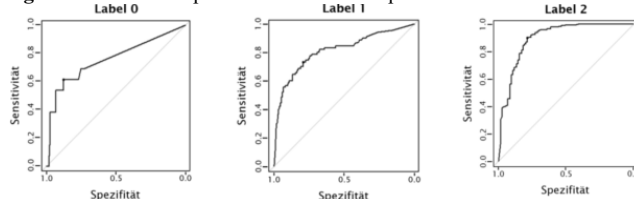


Figure 2. ROC curve per label from the comparison between N2 and ML.

The area below the ROC curve is a measure of the quality of the evaluation per label [11]. This means that if the ML algorithm cannot detect labels, the ROC curve forms a diagonal (50% area below the ROC curve). The better the ML algorithm, the larger the area below the ROC curve. The area below the ROC curve can also be seen as the probability that a label will actually be evaluated from the ML as the label from the neuropsychologist.

Table 2. Area below the ROC curve.

	N1 <-> ML	N2 <-> ML
Label 0	71.5%	75.6%
Label 1	72.0%	81.3%
Label 2	82.6%	89.7%
Average	75.4%	82.2%
Average total	78.8%	

This threshold value was used again to calculate the agreement in percent and in kappa. In addition, the number of drawings that must be checked manually for each threshold value was specified. This shows how the agreement changes and how many drawings have to be checked manually by a neuropsychologist.

With the threshold, the N1 and ML has an agreement of 80.3% (Kappa = 0.62) and the agreement between N2 and ML was 86.6% (Kappa = 0.73). N1 and N2 have to manually check 53 drawings (of the 305).

The significance test is intended to show how many ratings of the ML algorithm do not match with the ratings of N1 and N2 and are not submitted to them for a manually check ($\beta = 20\%$, error type 2) and how many ratings matched, but are submitted to the N1 and N2 for a manually check ($\alpha = 5\%$, error type 1). Follow hypothesis was defined for the significance test: H_0 : the ML rated the drawing equal to the neuropsychologist. H_1 : the ML rated the drawing not equal to the neuropsychologist.

Table 3. Significance test of N1 <-> ML (left) and N2 <-> ML (right).

	H ₀	H ₁		H ₀	H ₁
H ₀	63.0%	19.7%	H ₀	68.9%	13.4%
H ₁	9.5%	7.9%	H ₁	9.8%	7.9%

The error type 1 (cell H_1 / H_0) is on both significance tests over the 5% and not significant. The ML rated 9.5 / 9.8% of the drawings equal to N1 / N2, but with a reliability below the threshold. So those drawings have to be rated manually from a neuropsychologist.

The error type 2 (cell H_0 / H_1) should be as low as possible, because these drawings are not submitted to the neuropsychologist for a manually rating, although they are wrongly assessed. But as we can see, the error type 2 is in both cases under 20%, so β is significant.

Table 4. Overview comparison N1 with ML algorithm and N2 with ML algorithm with threshold of 0% and 78.8%.

	N1 <-> ML		N2 <-> ML	
Threshold (%)	0	78.8	0	78.8
Agreement (%)	72	80.3	79	86.6
Agreement (K)	0.45	0.62	0.56	0.73

4. Discussion

With our results is it possible to execute the BVMT-R Test in a digital way. The rating of the drawings is given by a trained ML algorithm either in a full automated way, or in

a semi-automated way with the possibility to set a threshold value for the reliability of the rating under which the drawing is manually rated by a neuropsychologist.

The biggest challenge to answer the research question was to determine the quality of the ratings received in the test data because there are no rated reference data sets (gold standard rating). With our solution, to measure the matches of the ratings of the test data performed by the COGITO GmbH with the ratings of the University Hospital Zurich, and find out it was 81%, we were then able to compare the results obtained in automated or semi-automated way with our ML algorithm with the results of either one of two neuropsychologists (N1 and N2), results with N1 showing the internal and N2 the external validation.. Consequently, to answer the research question positively the rating of the ML algorithm had to reach at least the same matching as the two professional raters.

With this result it is conceivable to create further projects in this area of science. A more detailed consideration could be a completion of the entire digital BICAMS test set using ML.

In summary, a semi-automated rating with the use of ML algorithms of patient-drawn drawing is possible. The deviation of this solution (matching 80.3% and 86.6% with the two neuropsychologists) is in the same range as the deviation between those two independent professionals (matching of 81%).

5. References

- [1] Broicher, Sarah Dinah. Neuropsychologie bei Multipler Sklerose [Internet]. Neuropsychologie bei Multipler Sklerose. 2014 [cited 2019 Feb 27]. Available from: <https://www.rosenfluh.ch/media/psychiatrie-neurologie/2014/05/Neuropsychologie.pdf>
- [2] Benedict, Ralph H. B. Brief Visuospatial Memory Test-Revised | BVMT-R [Internet]. [cited 2018 Dec 28]. Available from: <https://www.parinc.com/Products/Pkey/30>
- [3] Benedict RHB, Amato MP, Boringa J, Brochet B, Foley F, Fredrikson S, et al. Brief International Cognitive Assessment for MS (BICAMS): international standards for validation. *BMC Neurol*. 2012 Jul 16;12:55.
- [4] Birchmeier ME, Studer T. Automated Rating of Multiple Sclerosis Test Results Using a Convolutional Neural Network. *Studies in Health Technology and Informatics*. 2019;105–108.
- [5] Zhang W, Itoh K, Tanida J, Ichioka Y. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Applied Optics*. 1990 Nov 10;29(32):4790.
- [6] Aghdam HH, Heravi EJ. Guide to convolutional neural networks: a practical application to traffic-sign detection and classification [Internet]. 2017 [cited 2018 Oct 30]. Available from: <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1520793>
- [7] Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56(1):45–50.
- [8] Cohen J. Statistical power analysis for the behavioral sciences. [Internet]. 1988 [cited 2019 May 29]. Available from: <http://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf>
- [9] Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 1960 Apr 1;20(1):37–46.
- [10] Caneda MAG de, Cuervo DLM, Marinho NE, Vecino MCA de, Caneda MAG de, Cuervo DLM, et al. The Reliability of the Brief Visuospatial Memory Test - Revised in Brazilian multiple sclerosis patients. *Dementia & Neuropsychologia*. 2018 Jun;12(2):205–11.
- [11] Businger W, Bigler V. Medizinische Statistik [Internet]. Berne Fachhochschule - Technik und Informatik; 2019 [cited 2019 May 26]. Available from: <http://moodle.bfh.ch>

Discovering New Social Determinants of Health Concepts from Unstructured Data: Framework and Evaluation

Joao H. BETTENCOURT-SILVA^a, Natalia MULLIGAN^a,
Marco SBODIO^a, John SEGRAVE-DALY^b, Richard
WILLIAMS^c, Vanessa LOPEZ^a, Carlos ALZATE^a

^a *AI for Health & Social Care Research Group, IBM Research,
Ireland*

^b *IBM Watson Health, Dublin, Ireland*

^c *NIHR Greater Manchester Patient Safety Translational Research
Centre, University of Manchester, Manchester, UK*

Abstract. Social determinants of health (SDoH) are the complex set of circumstances in which individuals are born, or with which they live, that impact their health. Integrating SDoH into practice requires that information systems are able to identify SDoH-related concepts from charts and case notes through vocabularies or terminologies. Despite significant standardisation efforts across healthcare domains, SDoH coverage remains sparse in existing terminologies due to the broad spectrum of this domain, ranging from family relations, risk factors, to social programs and benefits, which are not consistently captured across administrative and clinical settings. This paper presents a framework to mine, evaluate and recommend new multidisciplinary concepts that relate to or impact the health and well-being of individuals using a word embedding model trained from a large dynamic corpus of unstructured data. Five key SDoH domains were selected and evaluated by domain experts. The concepts resulting from the trained model were matched against well-established meta-thesaurus UMLS and terminology SNOMED-CT and, overall, a significant proportion of concepts from a set of 10,000 candidates were not found (31% and 28% respectively). The results confirm both the gaps in current terminologies and the feasibility and impact of the methods presented in this paper for the incremental discovery and validation of new SDoH concepts together with domain experts. This sustainable approach facilitates the development and refinement of new and existing terminologies and, in turn, it allows systems such as Natural Language Processing (NLP) annotators to leverage SDoH concepts across integrated care settings.

Keywords. Social Determinants of Health, Terminologies, Vocabularies

1. Introduction

Throughout human societies, health and illness are not homogeneously distributed due to disparities in social and economic aspects. The aspects that underpin such disparities are commonly referred to as the Social Determinants of Health (SDoH) and examples include intangible factors such as “political, socioeconomic, and cultural constructs, as well as place-based conditions including accessible healthcare and education systems, safe environmental conditions, well-designed neighbourhoods, and availability of

healthful food” [1]. The significant impacts on health and costs associated with SDoH have been reported [2] and in order for social aspects to be best utilised for service evaluation or outcome assessment and improvement, this information should be routinely collected by organisations across the integrated care continuum. However, despite technological advances, SDoH information is neither routinely nor systematically collected in Electronic Health Records (EHRs) and lacks standardisation [3,4]. The US Institute of Medicine was among the first to identify SDoH domains and recommend they be routinely collected and made available in EHRs [3]. When available, SDoH information is often stored in natural language as part of unstructured case notes which further hinders efforts to identify cases and select cohorts. Terminologies and meta-thesaurus such as SNOMED-CT and UMLS can be used to encode health-related observations or activities into a canonical form. A recent exploratory analysis [5] has confirmed the limitations of current healthcare vocabularies in documenting SDoH-related information and has called for more comprehensive, coherent, and user-friendly SDoH code sets [6] to support and facilitate the rapidly evolving health care use cases[5]. Techniques for bridging clinical and social domains using word embeddings have been proposed to explore cross-domain spaces [7] and calls for action to create standards for representing SDoH data have been suggested [3]. Recent initiatives include Health Level Seven’s (HL7) Gravity Project [8] and LOINC’s models for the representation of screening assessments and measures of SDoH [9]. Healthcare terminologies have begun adopting concepts in social domains through consultation with expert groups but sustainable processes and methodologies to support these efforts are lacking.

This paper proposes a computational approach to facilitate the process of seeking new terminology concepts as well as refreshing existing terminologies. The framework proposed in this paper leverages existing word embedding techniques [10, 11, 12] in order to mine new candidate terms from large corpora of unstructured data. Other approaches have focused on developing ontologies and taxonomies [11] yet the work presented in this paper aims to incorporate expert humans-in-the-loop in order to recommend new concepts not included in existing terminologies. This approach can also be used to refine existing terminologies with categories for specific domains.

2. Methods and Experiments

The proposed framework takes a list of seeded terms as input and returns a list of relevant concepts with domain categories assigned. This process relies on training a model to discover relationships between terms in relevant corpora and is supported by a human-in-the-loop evaluation step to ascertain the relevance of the discovered terms. Framework steps are summarised in Figure 1 and described in detail in the next sections.

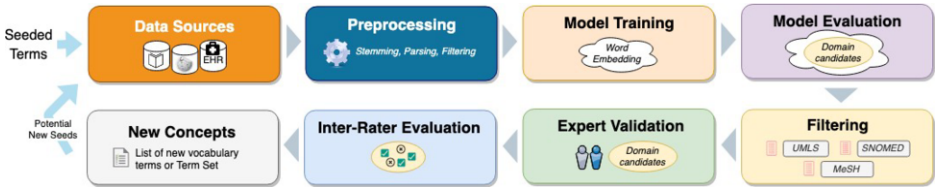


Figure 1. Proposed framework for recommending new terminology concepts.

Training a comprehensive SDoH model requires a list of relevant seeded term(s) for the first iteration of the process. Together with domain experts (DEs), 57 distinct seeded terms were selected for our experiments. The terms were extracted from relevant literature definitions of SDoH (WHO, Marmot & Wilkinson [13], CDC, Wikipedia).

2.1. Data Sources and Preprocessing

The first step consists of obtaining a corpus of unstructured data suitable for model training. A collection of case notes may be obtained from EHRs, case notes, books, or other sources of knowledge. Data sources should ensure coverage of the relevant domain areas in the seeded terms and this can be achieved in consultation with the DEs.

In order to train the comprehensive SDoH model presented in this paper, we prepared a corpus of text by crawling the set of Wikipedia pages hierarchically linked through DBpedia's taxonomy to the seeded terms. We curated the collected corpus by removing irrelevant pages, and obtained 2,134 distinct pages, with 3,282,508 non-unique terms. Finally, we pre-processed the collected corpus by removing metadata, images, and tables, and by applying lemmatisation and filtering of non-textual characters.

2.2. Model Training and Model Evaluation

This step focuses on the application of word embedding techniques on the prepared corpus of text. The purpose is to train a model able to capture contextual similarities between terms and term classes. The SDoH model was built using state-of-the-art embedding techniques to discover domain-specific n-grams from the corpus of text [10]. We trained a skip-gram model with negative sampling and subsampling of the frequent words in the training corpus. Experiments to tune the model hyperparameters were undertaken and found 100-dimension embeddings, computed with a context window of 5, and n-grams of size 3, provided a rich domain representation for our task. Other techniques such as pre-trained word embedding models may be used in this step yet these typically only yield unigrams. The resulting trained model contained 77,232 terms.

For evaluation, the DEs selected five terms each defining specific SDoH domains of interest: *Housing*, *Criminal Justice*, *Financial Services*, *Health Behaviours* and *Homeless*. For each of the five terms, a set of similar (candidate) terms was computed by selecting the closest 2000 neighbours in the model's high-dimensional space using Euclidean distance. The ranked set of 2000 candidate terms was considered large enough to be evaluated by DEs in a timely manner. The set was later randomised for the DEs.

2.3. Filtering and Expert Validation

In the filtering step, candidate terms from each of the five domains of interest were programmatically checked against UMLS and SNOMED-CT terminology. Each term was labelled as *matching* when either a partial or exact match was found in the terminology (via its API), or *no match*. It is possible to use more complex matching algorithms and to filter out terms that already exist in terminologies (e.g. synonyms), however, for the purpose of developing and evaluating the framework proposed in this paper, no filters were applied, and partial and exact matching were combined.

The Expert Validation step consists of gathering information from DEs about the candidate terms' relevance in their specific domains. This was achieved by carrying out

90 minute workshop sessions with social workers where at least 3 different DEs of the same specialty evaluated the same set by responding *accept*, *reject* or *pass* for each term.

2.4. Inter-Rater Evaluation and Recommendation of New Concepts

In this step the results are analysed in order to determine consensus among DEs (inter-rater agreement based on simple majority was used due to small number of DEs, however, other methods are possible). Decisions are made for all candidate terms and a curated list of recommended new concepts is the output of the framework (Figure 1). This list may serve to refine the input list of seeded terms in order to further build or refine models.

3. Results and Discussion

The framework proposed in this paper enabled the training of a model based on SDoH seeded terms. The model was evaluated on five key SDoH domains through workshops with DEs (each domain seen by 3 distinct DEs with the exception of *Housing*). Figure 2 summarises the DEs evaluation results for each domain and depicts a trend where lower ranked terms were accepted less often than those nearest to each of the five domain terms.

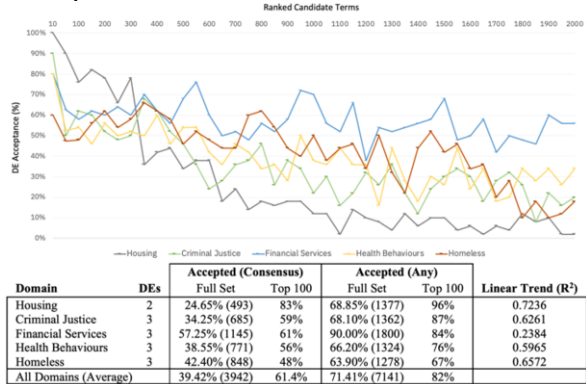


Figure 2. Proportion of terms accepted by DEs. For each domain, terms were ranked using Euclidean distance in the SDoH embedding model high-dimensional space.

Table 1 shows the results of matching each of the terms against SNOMED-CT and UMLS. Overall our findings show a large number of new terms that did not match either (on average 76% in SNOMED-CT and 83% in UMLS), with SNOMED-CT showing a better coverage throughout. Across all domains, on average, 40% of terms are accepted by DEs consensus of which 30% of them were not found in either (no match). *Financial Services* was the domain with most new relevant terms found across both SNOMED-CT (48%) and UMLS (50%) suggesting this area may be most lacking. *Housing* and *Homeless* are the two domains with the least number of new terms found and accepted by DEs (respectively 16% and 23% for SNOMED-CT, and 19% and 30% UMLS). Out of all 10,000 terms evaluated in this paper, the framework yielded 30% as concepts to be suggested for inclusion in terminologies and highlights terms from other terminologies of relevance to this domain. A limitation of the modelling approach is yielding less relevant n-gram combinations (e.g. ‘housing provides’ versus ‘housing programs’). However, the human-in-the-loop step helps filter out such cases. A subset of the concepts found could also be synonyms of existing terminology concepts and terms could appear

in both UMLS and SNOMED-CT. Since no partial or exact matches were found, however, it is still relevant to suggest bringing these terms in to terminologies.

Table 1. Results of matching terms against SNOMED-CT and UMLS.

Domain	SNOMED-CT				UMLS			
	Match % (N)		No Match % (N)		Match % (N)		No Match % (N)	
	Accepted	Rejected	Accepted	Rejected	Accepted	Rejected	Accepted	Rejected
Housing	8.50% (170)	25.75% (515)	16.15% (323)	49.60% (992)	6.05% (121)	17.80% (356)	18.60% (372)	57.55% (1151)
Criminal Justice	9.95% (199)	5.75% (115)	24.30% (486)	60.00% (1200)	7.55% (151)	3.75% (75)	26.70% (534)	62.00% (1240)
Financial Services	9.40% (188)	3.40% (68)	47.80% (956)	39.40% (788)	7.65% (153)	4.20% (84)	49.55% (991)	38.60% (772)
Health Behaviours	10.85% (217)	11.15% (223)	27.70% (554)	50.30% (1006)	6.80% (136)	7.50% (150)	31.75% (635)	53.95% (1079)
Homeless	19.30% (386)	17.25% (345)	23.05% (461)	40.40% (808)	12.75% (255)	10.90% (218)	29.60% (592)	46.75% (935)
Total (Average)	11.60%	12.66%	27.80%	47.94%	8.16%	8.83%	31.24%	51.77%
	24.26%		75.74%		16.99%		83.01%	

4. Conclusions

The lack of standardisation, terminologies and code sets in complex multidisciplinary domains remains a barrier to the successful documentation and use of SDoH in practice. This work assessed the coverage of existing terminologies in selected SDoH domains using expert humans-in-the-loop and proposed a framework for the discovery of SDoH concepts from unstructured data. Further work is needed to refine model training, to include additional categories for DEs to assign to terms and allowing new term/code sets to be defined in the process, and to inspect SDoH in other terminologies (e.g. ICD-11).

References

- [1] New England Journal of Medicine – Catalyst; 2017. Social Determinants of Health (SDOH). Retrieved from <https://catalyst.nejm.org/social-determinants-of-health>
- [2] McGovern L, Miller G, Hughes-Cromwick P. The relative contribution of multiple determinants to health. Health Affairs Health Policy Briefs. 2014 Aug 21;21.
- [3] Hripesak G, Forrest CB, Brennan PF, Stead WW. Informatics to support the IOM social and behavioral domains and measures. J Am Med Inform Assoc. 2015 Jul;22(4):921-4. doi: 10.1093/jamia/ocv035.
- [4] Adler NE, Stead WW. Patients in context--EHR capture of social and behavioral determinants of health. N Engl J Med. 2015 Feb 19;372(8):698-701. doi: 10.1056/NEJMp1413945.
- [5] Arons A, DeSilvey S, Fichtenberg C, Gottlieb L. Documenting social determinants of health-related clinical activities using standardized medical vocabularies, JAMIA Open. 2019. 2(1)81–88
- [6] Williams R, Kontopantelis E, Buchan I, Peek N. Clinical code set engineering for reusing EHR data for research: A review. J Biomed Inform. 2017 Jun;70:1-13. doi: 10.1016/j.jbi.2017.04.010.
- [7] Bettencourt-Silva J, Mulligan N, Cullen C, Kotoulas S. Bridging Clinical and Social Determinants of Health Using Unstructured Data. Stud Health Technol Inform. 2018; 255:70-74.
- [8] Health Level Seven, Gravity Project; 2019. Retrieved from <http://www.hl7.org/gravity>
- [9] Vreeman D. 2018; Advancing the interoperability of social and behavioral determinants of health. LOINC. Retrieved from <https://loinc.org/sdh/webinar>
- [10] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality. arXiv preprint arXiv:1310.4546. 2013 Oct 16.
- [11] Grefenstette G, Muchemi L. Determining the Characteristic Vocabulary for a Specialized Dictionary using Word2vec and a Directed Crawler. arXiv preprint arXiv:1605.09564. 2016 May.
- [12] Gu Y, Leroy G, Pettygrove S, Galindo MK, Kurzius-Spencer M. Optimizing Corpus Creation for Training Word Embedding in Low Resource Domains: A Case Study in Autism Spectrum Disorder (ASD). AMIA Annu Symp Proc. 2018 Dec 5
- [13] Marmot M, Wilkinson R. *Social Determinants of Health*, Oxford University Press, Oxford. 2005.

Distinguishing Septal Heart Defects from the Valvular Regurgitation Using Intelligent Phonocardiography

Arash GHAREHBAGHI^{a,1}, Amir A. SEPEHRI^b and Ankica BABIC^{c,d}

^a*Department of Innovation, Design and Technology, Mälardalen University, Sweden*

^b*CAPIS Biomedical Research and Development Centre, Mon, Belgium*

^c*Department of Biomedical Engineering, Linköping University, Sweden*

^d*Department of Information Science and Media Studies, University of Bergen, Norway*

Abstract

This paper presents an original machine learning method for extracting diagnostic medical information from heart sound recordings. The method is proposed to be integrated with an intelligent phonocardiography in order to enhance diagnostic value of this technology. The method is tailored to diagnose children with heart septal defects, the pathological condition which can bring irreversible and sometimes fatal consequences to the children. The study includes 115 children referrals to an university hospital, consisting of 6 groups of the individuals: atrial septal defects (10), healthy children with innocent murmur (25), healthy children without any murmur (25), mitral regurgitation (15), tricuspid regurgitation (15), and ventricular septal defect (25). The method is trained to detect the atrial or ventricular septal defects versus the rest of the groups. Accuracy/sensitivity and the structural risk of the method is estimated to be 91.6%/88.4% and 9.89%, using the repeated random sub sampling and the A-Test method, respectively.

Keywords. Time growing neural network, Intelligent phonocardiography, A-Test method, septal heart defects, heart sound signal

1. Introduction

Since the last few decades when extracting medical information from heart sound has become a topic of study, several machine learning methods have been proposed for learning details of the heart sound signals for the classification purposes [1][2][3]. Although deep learning methods initiated a considerable change in the processing methods sophisticated for medical signals including heart sound signal [4][5], several research questions should be answered in terms of both the theoretical and the applicative contents. PhonoCardioGram (PCG) is a device for recording of mechanical activity of heart, as reflected by the heart sound. The term Intelligent PhonoCardioGraphy that has recently appeared in this contexts, implies on a computerised PCG supported by the intelligent machine learning

¹Corresponding Author: Arash Gharehbaghi, Department of Innovation, Design and Technology, Mälardalen University, Västerås, Sweden; E-mail: arash.ghareh.baghi@mdh.se.

methods to extract medical information from the signal [6][7][8]. Time Growing Neural Network (TGNN) has been introduced as a powerful learning method either for classifying cyclic time series [4], or for extracting medical information from the non-cyclic stochastic time series [5], in parallel to the hybrid methods [9][10][11]. Effectiveness of TGNN has been investigated for symptom detection of PCG signals, and also for screening a certain heart abnormality among other possible pathologies [11][12][5]. An important applicative aspect of IPCG is its potential in patient prioritisation in terms of the disease severity. For example, a valvular regurgitation in children is managed differently from septal leakage, as the former might not effect growth of the children, whereas the later which can effect growth of the patient.

This paper presents a method for diagnosing a group of paediatric patients with septal defect from the other group of valvular leakage, the two pathological conditions manifested by systolic murmur. Similarity of the disease manifestations in auscultation make the diagnosis a complicated task, which necessitates expensive investigations even for those with mild defects. It is important to note that timely detection of septal defects can prevent further complications including pulmonary artery defect, ventricular hypertrophy and dilatation, pulmonary hypertension, and heart failure, each can put negative impact on the growth of the children suffering from, which is sometimes irreversible. The presented method employs TGNN for extracting medical information from a heart sound recording regarding the heart septal defects, showing an important application of medical informatics. The resulting technology can be easily incorporated into an IPCG to help practitioners or nurses at primary healthcare centres to detect and prioritise the patients for further investigation. The resulting IPCG can provide valuable medical information even for a paediatric cardiologist to be oriented in the echocardiographic investigation [13].

2. Material and Methods

2.1. Data Collection

Heart sound signals of 10 second duration were recorded from the children referrals to Tehran University of Medical Sciences, using a WelchAllyn Meditron Analyzer in conjunction with a portable computer with 44100 Hz of sampling frequency and 16 bit resolution. All the referrals or their legal guardians gave the informed consent for participation in the study, which was conducted according the Good Clinical Practice, and complied with the World Medical Association and Helsinki Declaration. The referrals were investigated by a paediatrician as well as paediatric cardiologist who used echocardiography as the gold standard along with complementary tests like chest X-Ray, in accordance to the guideline of Tehran University of Medical Sciences. Six groups of paediatric individuals were included in the study: healthy children with no audible murmur (NM), and with audible innocent murmur (IM), as well as abnormal children with Ventricular Septal Defect (VSD), Atrial Septal Defect (ASD), Mitral Regurgitation (MR) and Tricuspid Regurgitation (TR). The patient population is listed in Table 1.

Table 1. Study population.

Heart condition	Number of Patients	Age Range (years)	Average age \pm SD (years)
ASD	10	5 – 15	9.3 \pm 3.2
IM	25	2–14	6.7 \pm 3.7
NM	25	4 – 15	8.6 \pm 3.4
MR	15	4 – 18	11.8 \pm 4.1
TR	15	6 – 18	12.6 \pm 4.4
VSD	25	1 – 9	3.9 – 2.4

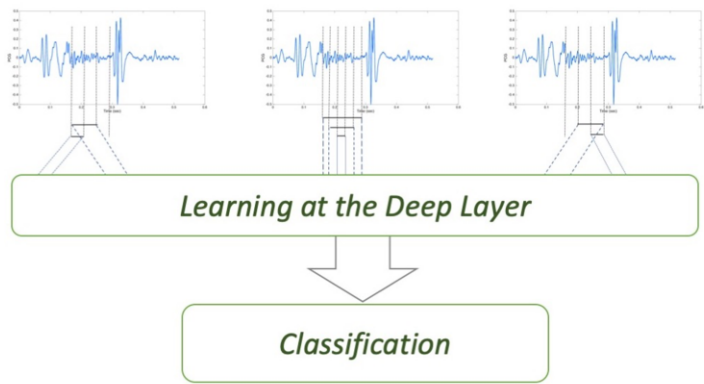


Figure 1. The flowchart of the learning method.

2.2. The processing algorithm

The signals are first down-sampled to 2 KHz using the antialiasing low-pass filter and then normalised. The processing method is based on our deep learning method, in which spectral contents of heart sound signal is calculated using three fashions of growing windows. In this technique, the systolic part of PCG signals are characterised by its spectral contents calculated over three different schemes of temporal windows: the forward, the backward and the mid-growing windows. Details of the growing method can be found in [4]. The number of the growing windows for each scheme is selected to be 3 based on the medical considerations. Figure 1 illustrates the growing time schemes. We used Fisher criteria for finding the most discriminative frequency bands for spectral calculation. This is performed at the deep learning layer in which the most discriminative frequency band is found for each temporal window, and the spectral energies are used by a three layer perceptron neural network with a hidden/output layer of 10/1 neurone for the ultimate learning process. The neural network is trained using back propagation error method in which the output layer performs the binary classification of septal defect against the rest of the classes .

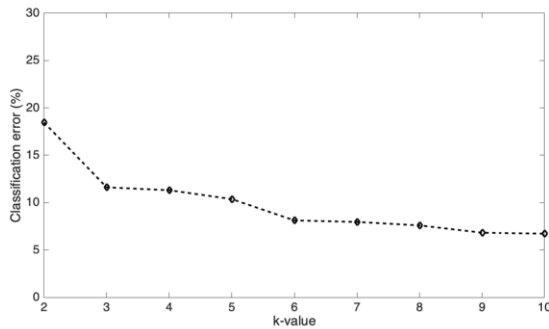


Figure 2. Results of the A-Test method for different index of k -value.

2.3. The validation method

The method is validated by using repeated random sub-sampling method as well as the A-Test for the structural risk evaluation. In the repeated random sub-sampling, 30% of data is randomly selected for testing and the rest for training the method, and the two performance measures, accuracy and sensitivity, are calculated, where the former is defined the percentage of the correctly classified samples whereas the later which is defined as the percentage of the correctly classified samples from the patient group (those patients with septal defect). This procedure is repeated several times and the statistical descriptive of the performance measures are found. The A-Test method is based on using K-fold validation method with different values of K , spanning from 2 to half of the minimum group size [4].

3. Results

The repeated random sub-sampling with 100 iterations was applied to validate performance of the method. Average of the accuracy and the sensitivity is estimated to be 88.4 and 91.6 with and standard deviation of ± 3.9 and ± 5.7 , respectively. In order to evaluate structural risk of the method, the A-Test method is employed. The average classification error is estimated to be 9.89% using the A-Test method. Figure 2 shows results of the A-Test method, exhibiting a decreasing trend for the classification error, which implies on a good capacity of enhancement with larger training data. As can be seen, the method performance is by far better than a paediatric cardiologist who relies on the conventional auscultation, as the studies show the average screening accuracy is less than 80% [3].

4. Discussion

This study suggested a method for distinguishing two important cardiac abnormalities, the septal defects and valvular regurgitation. Ventricular septal defects must be diagnosed at the early ages, and the diagnosed patients should undergo appropriate disease managements, in contrast with the valvular regurgitation which are found at the later ages of the childhood. An important aspect of this study is the use of sophisticated time grow-

ing neural network for characterising heart sound signal, and hence extracting important medical information from the signals. This innovative method allows the IPCG to provide a more comprehensive information to practitioners or family doctors at primary healthcare centres, resulting in a more efficient patient prioritisation. This avails access to paediatric cardiologists to those of rather urgent need sooner than those of need to a regular supervision, and also improves automatic extraction of medical information from IPCG.

Acknowledgement

The authors would like to thank Prof. A. Kocharian for his valuable cooperation in data collection. This study was supported by the CAPIS Inc., Mons, Belgium, and also by the KKS financed research profile in embedded sensor systems for health at Mälardalen University, Västerås, Sweden.

References

- [1] I. Cathers, Neural network assisted cardiac auscultation, *Artificial Intelligence in Medicine* 7 (1995) 53–66.
- [2] C. G. DeGroff, S. Bhatkar, J. Hertzberg, R. Shandas, L. Valdes-Cruz, R. L. Mahajan, Artificial neural network-based method of screening heart murmurs in children, *Circulation* 103 (2001) 2711–2716.
- [3] R. L. Watrous, W. R. Thompson, S. J. Ackerman, The impact of computer-assisted auscultation on physician referrals of asymptomatic patient with heart murmurs, *Clin. Cardiol.* 31 (2008) 79–83.
- [4] A. Gharehbaghi, M. Linden, A deep machine learning method for classifying cyclic time series of biological signals using time-growing neural network, *IEEE Transactions on Neural Networks and Learning Systems* 29 (2018) 4102–4115.
- [5] Arash Gharehbaghi, Maria Linden, Ankica Babic, An artificial intelligent-based model for detecting systolic pathological patterns of phonocardiogram based on time-growing neural network, *Applied Soft Computing* 83 (2019) 105615.
- [6] A. Sepehri, Amir, A. Kocharian, A. Janani, A. Gharehbaghi, An intelligent phonocardiography for automated screening of paediatric heart diseases, *Journal of Medical Systems* 40 (2015).
- [7] Arash Gharehbaghi, Amir A Sepehri, Armen Kocharian, Maria Linden, An intelligent method for discrimination between aortic and pulmonary stenosis using phonocardiogram, *World Congress on Medical Physics and Biomedical Engineering*, June 7-12, 2015, Toronto, Canada, 1010–1013.
- [8] Arash Gharehbaghi, Amir A Sepehri, Maria Linden, Ankica Babic, Intelligent phonocardiography for screening ventricular septal defect using time growing neural network, *Informatics Empowers Healthcare Transformation* 238 (2017) 108-111.
- [9] Arash Gharehbaghi, Thierry Dutoit, Amir A Sepehri, Armen Kocharian, Maria Linden, A novel method for screening children with isolated bicuspid aortic valve, *Cardiovascular engineering and technology* 6 (2015) 546-556.
- [10] Arash Gharehbaghi, Per Ask, Eva Nylander, Birgitta Janerot-Sjoberg, Inger Ekman, Maria Linden, Ankica Babic, A hybrid model for diagnosing sever aortic stenosis in asymptomatic patients using phonocardiogram, *World Congress on Medical Physics and Biomedical Engineering*, June 7-12, 2015, Toronto, Canada, 1006–1009.
- [11] Arash Gharehbaghi, Amir A Sepehri, Maria Linden, Ankica Babic, A Hybrid Machine Learning Method for Detecting Cardiac Ejection Murmurs, *EMBECE & NBC 2017*, 787–790.
- [12] A Gharehbaghi, A Babic, Structural Risk Evaluation of a Deep Neural Network and a Markov Model in Extracting Medical Information from Phonocardiography, *Studies in health technology and informatics* 251 (2018) 157–160.
- [13] S.M. Debbal, F.Bereksi-Reguig, Time-frequency analysis of the first and the second heartbeat sounds, *Applied Mathematics and Computation* 184 (2007) 1041–1052.

Document Oriented Graphical Analysis and Prediction

Shorabuddin SYED, MS^{a,1}, Mahanazuddin SYED, MS^a, Hafsa Bareen SYEDA, MD^a, Fred PRIOR, PhD^a, Meredith ZOZUS, PhD^b, Melody L. PENNING, PhD^a, Mohammed ORLOFF, PhD^{c,d}

^aDepartment of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA ^bDepartment of Population Health Sciences, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA ^cDepartment of Epidemiology, University of Arkansas for Medical Sciences, Little Rock, AR, USA

^dWinthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Abstract. In general, small-mid size research laboratories struggle with managing clinical and secondary datasets. In addition, faster dissemination, correlation and prediction of information from available datasets is always a bottleneck. To address these challenges, we have developed a novel approach, Document Oriented Graphical Analysis and Prediction (DO-GAP), a hybrid tool, merging strengths of Not only SQL (NoSQL) document oriented and graph databases. DO-GAP provides flexible and simple data integration mechanism using document database, data visualization and knowledge discovery with graph database. We demonstrate how the proposed tool (DO-GAP) can integrate data from heterogeneous sources such as Genomic lab findings, clinical data from Electronic Health Record (EHR) systems and provide simple querying mechanism. Application of DO-GAP can be extended to other diverse clinical studies such as supporting or identifying weakness of clinical diagnosis in comparison to molecular genetic analysis.

Keywords. Data integration, clinical trials, data quality, data processing

1. Introduction

Traditionally, to develop a relational database or data repository, first, the data requirement is analyzed and then data structure is designed based on the requirements [1]. The data structure remains rigid and can only hold the data elements that it was designed for [2]. Then, data integration is performed by feeding-in data from various sources into a data warehouse and processing it afterwards. Fixed source to target Extract Transformation and Load (ETL) processes are used to load data into a data repository [3]. Drawback of using relational databases is that it is difficult to keep up with changing data requirements and it takes months to modify source to target mapping and implement ETLs. Not only SQL (NoSQL) databases were developed to overcome fixed data structure limitation of relational data models [4].

¹ Corresponding Author, Shorabuddin Syed, University of Arkansas for Medical Sciences, Little Rock, AR, USA; E-mail: ssyed@uams.edu.

The schema-less design of NoSQL allows flexibility and scalability to add information to any field without a need to define the structure in advance compared to SQL databases [5-7]. One of the common NoSQL database is document-oriented database that uses documents as the structure for data storage and queries, document is a block of XML or JSON [7, 8]. Each document can have the same or different structure. MongoDB is one of the document-oriented database system, it is highly flexible, partition tolerant and consistent [4, 9].

Graph database models their schema as graphs, using nodes and edges. Each node represents an entity (patient, physician etc.) and they are connected by edges, representing relationship between the nodes[10]. Graph database is queried by traversing through the connected links. The graph database are expressive and can be easily conceptualized making them an ideal choice to adopt for data analysis[10]. Moreover, “graph databases excel in traversal-type queries ”[11], they facilitate deep knowledge discoveries using simple-to-mid complex queries. Neo4j (www.neo4j.com) is a graph database designed to manage information with graph-like nature.

DO-GAP attempts to take advantages of both the models (Document and Graph DB) and solves challenges (i) flexible schema that enables faster and more iterative development, and scalable databases (ii) clean and readily available data for graph database. DO-GAP uses MongoDB as document-based data repository and Neo4j graph database for simple analysis and easy visualization.

2. Methods

To address our hypothesis, a pipeline was built for processing of the incoming data, store it in primary data repository, cleansing and transferring study specific data to graph database for analysis and knowledge discovery as shown in Figure 1.

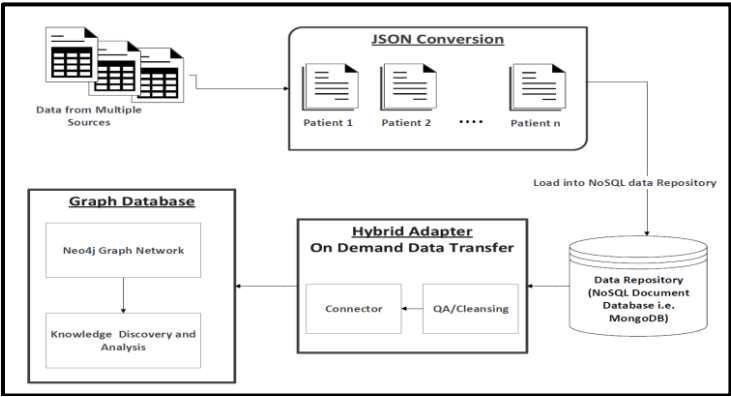


Figure 1. DO-GAP Architecture.

2.1. Data Architecture

The flow starts by collecting and assembling data from multiple sources, extract patient’s longitudinal data from the source files and create a JSON document for each patient. The

JSON files created will be merged into primary data repository, MongoDB. On-demand, study-specific data from MongoDB is cleansed and transferred to graph database, Neo4j. Laboratories may receive clinical and biological data from multiple sources in different format, frequency and file types. Metadata, a mapping document, for each source file is created to standardize the source data elements. JSON conversion tool creates a JSON document based on source and metadata files. The conversion is independent of source file format, thereby minimizing data pre-processing time. Figure 2, shows an example of patient demographic Comma Separated Values (CSV) file and its equivalent JSON file created using the JSON conversion tool. A master metadata file is maintained during the process.

Example Patient demographics file
id,full_name,age,Gender,ethnicity,DOB,DOD,alive_Dead,Zipcode,mrn,race 109563,Jon Doe,68,Male,Non-Hispanic or Latino,5/1/1951,NULL,Y,72207,M000501,White
Example standardization Patient demographics file
{ "patient_id":109563,"patient_name": "Jon Doe","age":68,"sex": "Male","ethnicity": "Non-Hispanic or Latino","birth_date": "1/1/1990","death_date": null,"is_alive_yn": "Y","zip": 72207,"mrn": "M000501","race": "White" }

Figure 2. A sample demographic file and its equivalent standardized JSON file.

The JSON documents created are loaded into NoSQL document database i.e. MongoDB, using its native data import tool. To inform researchers on the available data fields in DO-GAP, a list of all field names (i.e. Keys) present in the MongoDB and their description is available. This list is updated when new data elements are imported into the database. Researchers can select from the available fields for data extraction and analysis.

Based on user selection, Hybrid adapter cleanses the selected data elements. The cleansing process includes applying standard rules such as removing potential duplicates and null values etc. This cleaned data set will be fed to Connector, it unwinds JSON documents and stores it as deconstructed array document using \$unwind, \$project, \$out MongoDB operators[4]. The deconstructed array document is then written to CSV file using mongoexport utility. Connector uses the generated CSV files to create nodes in Graph database i.e. Neo4j, and builds relationship between the nodes. By this approach only required data elements for a study is selected on-demand, cleansed and made available as a new study in graph database. Thereby overcoming Graph database’s data quality and volume limitations[12]. In Graph database component, the data extracted from MongoDB is loaded into Neo4j and is analyzed using native tools such as Neo4j browser and Bloom [13].

3. Application

About 2,300 patients with any type of lung cancer disease[14] were chosen for the study. Selected patient’s clinical facts, genomic profiles, comorbidities, social and family history were loaded into DO-GAP for knowledge discovery, resulting in 2,760 nodes and 9,683 relationships. We ran a Cypher query to find the common gene alterations in our sample. As shown in Figure 3, most patients have gene TP53 and LRP1B, TP53 and SPTA1 alterations. Cypher queries are simple to learn as non-technical users can query the database by conceptualizing the graphs, which is difficult with relational databases.

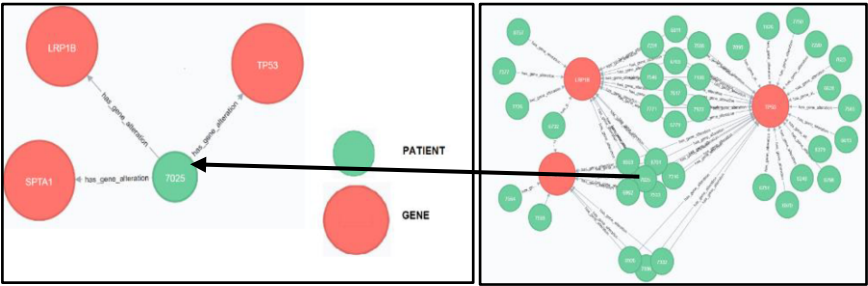


Figure 3. Partial graph showing links between Patient and their Gene alteration.

The output of graph analysis heavily depends upon connectivity between nodes [15], in DO-GAP we can click on any two gene nodes and can ask for common patterns (clinical facts) between them. Graph traverse through all links and displays every node (clinical facts) that is connected to these two gene nodes, hence the data or knowledge discovery.

Graph database can be queried using native visualization tools of Neo4j. In our case, we have used Bloom [16] to demonstrate the ease of querying. As shown in Figure 4, a simple English-like querying language syntax is used to find all patients with any gene alterations and are smokers. In addition, it displays the profile of query results, for example in Figure 4; we have 43 patients that are smokers with at most 10 gene alterations. Graph database query can be expanded to include other clinical facts along with gene information.

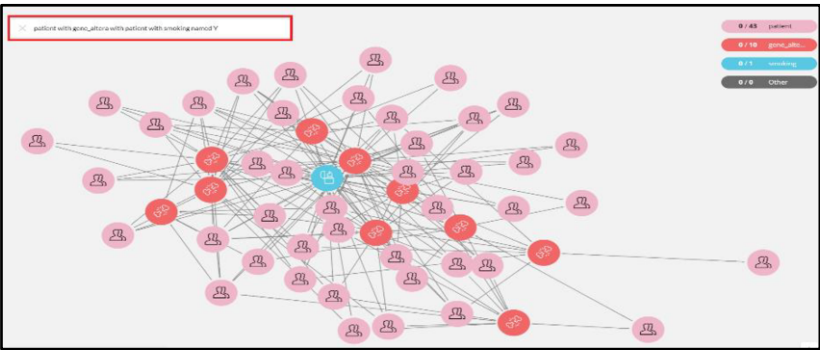


Figure 4. Very simple query to find patients who are smokers and has gene alteration.

4. Discussion and Conclusion

DO-GAP offers hybrid and flexible data integration, expansion, productive data analysis and prediction with minimal coding. Advantage of using DO-GAP is twofold (1) flexible schema that enable faster and more iterative development, and scalable NoSQL databases. (2) As strength of graph data analysis heavily depends on quality of data, we have achieved it by using hybrid adapter that automatically cleanse user selected data and import to graph database. Graph databases reduces query complexity compared to

both relational and MongoDB databases. Use of graph database reduces data analysis and data inference time; we have demonstrated this by two basic use cases (i) common gene alteration, and (ii) simplicity of querying heterogeneous data. As this was a pilot study, we selected the use cases that profile the data, which is the first step in research data analysis.

There are several limitations in this study that we wish to acknowledge. First, to resolve identity of individuals, i.e. entity resolution, which is a complex process in itself and is beyond the scope of this paper. Currently, DO-GAP process these records as separate subjects, in future, an entity resolution systems such as OYSTER [17], can be employed to resolve this limitation. Second, the cleansing component of hybrid adapter only checks for null values and duplicate clinical facts, but can be expanded to incorporate in-depth data quality checks. Finally, results of selected use cases can be easily reproduced using relational databases by technical users. However, DO-GAP was designed to be a data management and analysis tool for non-programmers such as lab technicians, for whom conceptualizing relational database schema and building appropriate queries is challenging task. This pilot project, paves way for future work that will seek to address in-depth graph queries for better findings and correlations. We envision to address aforementioned limitations in our future work that will empower users to explore critical insights which is currently not available.

Funding Acknowledgment and Consent: Patients data used were obtained under IRB approval (IRB#202880) at the University of Arkansas for Medical Sciences (UAMS). The study was supported partly by the Seeds of Science Grant by the Winthrop P. Rockefeller Cancer Institute.

References

- [1] Sahatqija K, Ajdari J, Zenuni X, Raufi B, Ismaili F. Comparison between relational and NoSQL databases 2018. 0216-21 p.
- [2] Kent WJCA. A simple guide to five normal forms in relational database theory. 1983;26(2):120-5.
- [3] H Inmon W. Building the Data Warehouse 1996.
- [4] David Hows Peter M, Eelco Plugge, Tim Hawkins. The Definitive Guide to MongoDB Second ed: Apress; 2013.
- [5] Schulz WL, Nelson BG, Felker DK, Durant TJS, Torres R. Evaluation of relational and NoSQL database architectures to manage genomic annotations. Journal of biomedical informatics. 2016;64:288-95.
- [6] Leavitt N. Will NoSQL Databases Live Up to Their Promise? Computer. 2010;43(2):12-4.
- [7] Wu T-LS. An Overview of Present NoSQL Solutions and Features. .
- [8] Imam AA, Basri S, Ahmad R, Watada J, González-Aparicio MT. Automatic schema suggestion model for NoSQL document-stores databases. Journal of Big Data. 2018;5(1):46.
- [9] Moniruzzaman ABM, Hossain S. NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison 2013.
- [10] Renzo Angles CG. Survey of graph database models. ACM Comput Surv. 2008.
- [11] Lysenko A, Roznovát IA, Saqi M, Mazein A, Rawlings CJ, Auffray C. Representing and querying disease networks using graph databases. BioData Min. 2016;9:23-.
- [12] Pokorný J, editor Graph Databases: Their Power and Limitations. Computer Information Systems and Industrial Management; 2015 2015//; Cham: Springer International Publishing.
- [13] NeoTechnology. <https://neo4j.com/bloom/>.
- [14] ICD10Data.com. Malignant neoplasm of bronchus and lung C34 [Internet]. [cited on 2018 Dec 9]. <https://www.icd10data.com/ICD10CM/Codes/C00-D49/C30-C39/C34->.
- [15] Rahman MR. Knowledge Base Population based on Entity Graph Analysis: Université Paris-Saclay; 2018.
- [16] NeoTechnology. Neo4j Bloom Quick Start. [Internet]. [cited on 2018 Dec 9]. Available from: <https://neo4j.com/docs/bloom-user-guide/current/about-bloom/>. [
- [17] Talburt JR, Zhou Y. A practical guide to entity resolution with OYSTER. 2013. p. 235-70.

Electronic Whiteboard in the Inpatient Care Management: Nurses' Perceptions About the Use

Liliana GIRALDO^{a,1} Bibiana SCHACHNER^a, Daniel LUNA^a, Sonia BENITEZ^a

^aDepartment of Health Informatics, Hospital Italiano de Buenos Aires

Abstract. In the hospital setting, the traditional method of information for nurses is to consult the Electronic Health Records (EHR) on several occasions to check the continuous updates of the prescriptions. This process, in practice, involves time and multiple setbacks such as: availability of equipment (there is no computer per nurse in the office), number of patients assigned (it involves entering each EHR one by one). This is how the implementation of the tool known as electronic whiteboard (e-whiteboard) came about. The aim of this study was to explore perceptions of nurses after implementation of e-whiteboard, recently introduced at Hospital Italiano de Buenos Aires. A Qualitative study, which consisted of 5 group interviews with a total of 27 participants. A convenience sampling was used (selection of nurses from different shifts), corresponding to 4 sectors of adult patient admissions and 1 sector of pediatric patients. The main findings were positive aspects such as: it was a simple, intuitive and easy to navigate interface, it allowed communication with other health professionals (physicians, nutritionists and kinesiotherapists), and it facilitated the visualization of real time information that favors the fulfillment of tasks in scheduled time. This tool is considered of great clinical importance for coordinated patient care because the visibility and availability of clinical information improves communication between professionals, streamlines workflow, and consequently accelerates shared decision making.

Keywords. Perceptions, nursing, electronic whiteboard, inpatient

1. Introduction

Advances in health have led organizations to incorporate technologies for patient care planning [1]. In the hospital setting, during the hospital stay, nursing work regarding patient care planning requires continuous updates. The traditional method, which consists of accessing the Electronic Health Records (EHR) involves multiple setbacks [2] such as: availability of equipment (there is not one computer per nurse in the office), or the number of patients assigned (it implies gaining access to every EHR one by one). Unlike EHR, electronic whiteboards (e-whiteboards) are being used as a tool to monitor [3,4] and improve the quality and safety of care [4].

Previous research has shown that e-whiteboards have certain advantages over traditional dry erase white-boards [5,6] (which have been and continue to be used in

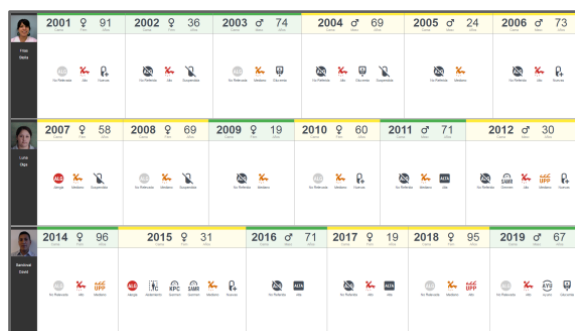
¹ Corresponding Author, Liliana. Giraldo. BSN. Department of Health Informatics, Hospital Italiano de Buenos Aires, Perón 4190, Buenos Aires (1199), Argentina. Email: liliana.giraldo@hospitalitaliano.org.ar

our local environment). These advantages include automated updating of patient information, the ability to visualize information in real time, and integration with other clinical information systems, and reduction of the workload of nurses [2,7]. While the use of tablets and mobile solutions in hospitals is currently being discussed, in our environment we are using more precarious tools and are just introducing e-whiteboards [4]. The aim of this study was to explore perceptions of nurses after implementation of e-whiteboard, recently introduced.

2. Methods

The Hospital Italiano de Buenos Aires (HIBA) is an institution with more than 1200 nurses and 750 beds in Buenos Aires, Argentina. The e-whiteboards, large electronic screens located on the nurse's station wall and out of sight of patients and visitors, were implemented in October 2017. This implementation were restricted to the adult and pediatric inpatient wards.

The e-whiteboard was developed with a user-centred design methodology, with a system that makes "retrieval information" in real time, addressed to nursing care staff. Categorization of information are displayed in around 84 icons in a horizontal distribution. The categories are: hospitalization data (patient's patronymic data, bed number, gender, age), alerts (for example: new prescriptions and/or changes in medication, isolation status, allergies), risks (pain scale assessment, pressure ulcers) and tasks (laboratory requests, patient transfers). The system interacts with the data from the admission, discharge and transfer (ADT) system, operating room agenda, EHRs, medical prescriptions and Computerised Physician Order Entry (CPOE) for complementary tests, nursing e-chart. (See Figure 1).



2001 ♀ 91	2002 ♀ 36	2003 ♂ 74	2004 ♂ 69	2005 ♂ 24	2006 ♂ 73
2007 ♀ 58	2008 ♀ 69	2009 ♀ 19	2010 ♀ 60	2011 ♂ 71	2012 ♂ 30
2014 ♀ 90	2015 ♀ 31	2016 ♂ 71	2017 ♀ 19	2018 ♀ 95	2019 ♂ 67

Figure 1: e-whiteboard screen

2.1 Design

A qualitative study to explore perceptions of nurses after implementation of e-whiteboard [8] was performed using group interviews. The data obtained, were homogeneous in terms of the institutional hierarchical position, there were not an inhibition of free expression by power relations and / or hierarchy, and it was achieved a greatest diversity of possible opinions.

A convenience sampling was carried out that included the selection of nurses from different shifts (morning, afternoon, evening and Saturday, Sunday or holiday). These Nursing staff belonged to four adult wards and one pediatric ward. From March to April 2018, 5 group interviews were conducted, with a duration of 90 minutes each one following an open-ended questionnaire. Grounded theory was used to analyze the transcribed data and build convergent themes. The relevance and importance of themes was constructed by examining frequency, convergence, and intensity. The presentation of results is oriented by these categories.

3. Results

A total of 27 nurses, with no age limit, from both genders, and from the general hospitalization area, participated in different shifts. Those nurses who carried out administrative management tasks (not patient care) were excluded. Six dimensions emerged from this study: use, communication, workflow, data quality, training and technical support. Each dimension is detailed with its respective findings (See Table 1).

Table 1. Dimensions and findings

Dimension	Description	Example
Dimension one: Use	Two components were identified: interface design and information availability. Nurses' perceptions were positive in relation to a simple and intuitive interface, simple navigation based on user-centered design guidelines, and the interaction with other systems of the institution. Likewise, it was highlighted that the iconography used facilitates its interpretation and use. In this sense, the most used icons were: assessment of pain scale, fall risk, isolation status, early warning, diet and the status of pharmacological prescriptions (new, modified, and suspended).	<i>"I use the e-whiteboard to visualize the pending tasks and for patient care planning"</i> <i>"It's very complete, because it shows everything you need to plan patient's care"</i>
Dimension two: Communication	Four structural elements of communication were determined using group interviews: the e-whiteboard properties, its physical location, permanent visualization, and the health personnel that must interact with the e-whiteboard (nurses, nutritionists, kinesiologists, stretchers, physicians, among others). Although they expressed positive perceptions related to the communication established with other health professionals (physicians, nutritionists and kinesiologists), it was not useful for the communication between them (nursing staff) to perform the shift pass.	<i>"It is very useful for communication with nutritionists and kinesiologists who come to evaluate patients"</i> <i>"Allows easy identification of the nurse in charge of the patient"</i>
Dimension Three: Workflow	A perception of a workflow improved was identified, related to: decrease of constant EHR access, visualization of relevant information and pending tasks in real time, and facilitates fulfilling tasks in scheduled time (without delays, improving timely patient care).	<i>"When there are changes in medical indications, they are immediately visualized, this helps to better plan our routine"</i>
Dimension Four: Data	The interviewees expressed positive aspects related to: the accuracy, reliability and quality of the information	<i>"It is reliable and secure data that are displayed on</i>

quality	displayed. They emphasized the capability to carry out patient care planning (due to the accessible and timely information), without generating complex processes to understand and assimilate the meaning of the data through iconography and colors.	<i>the e-whiteboard"</i>
Dimension Five: Training	The nurses mentioned as a positive aspect, the brochures provided with the information of the e-whiteboard. However, they consider negative the lack of training or follow-up during the implementation.	<i>"The information displayed is easily understood, but I would have liked accompaniment in my hospital sector"</i>
Dimension Six: Support	They mentioned as a negative aspect the lack of a help desk support that would allow the resolution of problems related to the e-whiteboard, and difficulties when the board failed.	<i>"When a e-whiteboard failed, we didn't know what to do"</i>

4. Discussion

This research proposed to explore the perceptions of nurses in relation to the implementation of the e-whiteboard. The main findings were positive aspects such as: the e-whiteboard was a simple, intuitive and easy to navigate, allowed communication with other health professionals (physicians, nutritionists and kinesiologists), and facilitated the real time visualization of information that favors the fulfillment of tasks in a scheduled time. The most used icons were: pain scale assessment, fall risk, isolation, early warning, diet and the status of pharmacological prescriptions.

This system is considered important for a coordinated patient care because the visibility and availability of clinical information improves communication between professionals, streamlines workflow [8], and consequently speeds up shared decision-making. Consistent with other studies [3,9], the e-whiteboard represents a technology that improves communication and coordinated patient care. On the other hand, the interviewees expressed certain negative aspects such as: it was not useful for communication between nursing staff to carry out the shift pass, the lack of training or coaching in its use, or the lack of a convenient support. In view of these findings, training and information dissemination should be planned in order to proceed in case of failure or to mitigate the risk.

The e-whiteboard could not be incorporated in the nursing pass workflow due to its physical location in the nursing office. The system could be viewed from the corridors by family members. Because information that must be hidden to maintain patient confidentiality (for example: sex, gender and other important clinical history), and after a consultation with the legal department, it was decided to use the bed numbers instead of the patient's name or initials.

Despite this is a unicentric study, the user-centered design methodology used facilitated the resolution of some specific needs of the end user, providing satisfaction and a good experience [10]. Post-implementation, the acceptance of the system by the staff is decisive [11], which has a direct impact on its use.

We believe that our findings can provide an approach to the understanding the complex factors related to the implementation of health technology, looking at staffs needs and concerns that are still present post-implementation [12], and which will be

worked on in future studies. This information is useful and can be expanded through a multi-hospital study.

Acknowledgements

This research was conducted under the Health Informatics Residency Program at HIBA. We want to thank to Dra. María Florencia Grande, for his advice and support, the Nursing Department and the nursing staff that participated.

References

- [1] Randell R, Greenhalgh J, Wyatt J, Gardner P, Pearman A, Honey S, and Dowding D, Electronic whiteboards: review of the literature, *Stud. Health Technol. Inform.* 210 (2015) 389–393.
- [2] Riley R, Forsyth R, Manias E, and Iedema R, Whiteboards: Mediating professional tensions in clinical practice, *Communication Medicine.* 4 (2007) 165–175. doi:10.1515/cam.2007.020.
- [3] Wong HJ, Caesar M, Bandali S, Agnew J, and Abrams H, Electronic inpatient whiteboards: Improving multidisciplinary communication and coordination of care, *International Journal of Medical Informatics.* 78 (2009) 239–247. doi:10.1016/j.ijmedinf.2008.07.012.
- [4] Gjære EA, and Lillebo B, Designing privacy-friendly digital whiteboards for mediation of clinical progress, *BMC Medical Informatics and Decision Making.* 14 (2014). doi:10.1186/1472-6947-14-27.
- [5] Wears RL, Perry SJ, Shapiro M, Beach C, Croskerry P, and Behara R, A Comparison of Manual and Electronic Status Boards in the Emergency Department: What's Gained and What's Lost?, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* 47 (2003) 1415–1419. doi:10.1177/154193120304701208.
- [6] Xiao Y, Kim YJ, Gardner SD, Faraj S, and MacKenzie CF, Communication technology in trauma centers: A national survey, *The Journal of Emergency Medicine.* 30 (2006) 21–28. doi:10.1016/j.jemermed.2005.04.014.
- [7] Egan M, Clinical dashboards: impact on workflow, care quality, and patient safety, *Crit. Care Nurs. Q.* 29 (2006) 354–361.
- [8] ED improves on already impressive wait times, *ED Manag.* 22 (2010) 6–7.
- [9] Bardram J, Bunde-Pedersen J, and Soegaard M, Support for activity-based computing in a personal computing operating system, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '06.* (2006). doi:10.1145/1124772.1124805.
- [10] Tramullas J, Documentos y servicios digitales: de la usabilidad al diseño centrado en el usuario, *El Profesional de La Informacion.* 12 (2003) 107–110.
- [11] Hwindingwi PP, *The Implementation of the Electronic Whiteboard in a Private Hospital: Challenges Experienced by Nurses in Gauteng.* 2013.
- [12] de Veer AJE, and Francke AL, Attitudes of nursing staff towards electronic patient records: A questionnaire survey, *Int. J. Nurs. Stud.* 47 (2010) 846–854.

Efficient Protection of Health Data from Sensitive Attribute Disclosure

Raffael BILD^{a,1}, Johanna EICHER^a and Fabian PRASSER^{b,c}

^aUniversity hospital rechts der Isar, Technical University of Munich, Germany

^bCharité - Universitätsmedizin Berlin, Berlin, Germany

^cBerlin Institute of Health (BIH), Berlin, Germany

Abstract. Biomedical research has become data-driven. To create the required big datasets, health data needs to be shared or reused out of the context of its initial purpose. This leads to significant privacy challenges. Data anonymization is an important protection method where data is transformed such that privacy guarantees can be provided according to formal models. For applications in practice, anonymization methods need to be integrated into scalable and robust tools. In this work, we focus on the problem of scalability.

Protecting biomedical data from inference attacks is challenging, in particular for numeric data. An important privacy model in this context is t -closeness, which has also been defined for attribute values which are totally ordered. However, directly implementing a scalable algorithmic representation of the mathematical definition of the model proves difficult. In this paper we therefore present a series of optimizations that can be used to achieve efficiency in production use. An experimental evaluation shows that our approach reduces execution times of anonymization processes involving t -closeness by up to a factor of two.

Keywords. data protection, anonymization, inference attacks, scalability

1. Introduction

Biomedical research, e.g. in the field of precision medicine which tailors healthcare to characteristics of individuals, is increasingly data-driven and leveraging methods from data science such as machine learning [1]. However, when creating the required big datasets, stringent privacy protection is mandated by laws and regulations. Hence, a wide range of safeguards has to be applied, including organizational and technical measures.

Data anonymization is an important building block for implementing technical privacy protection. The basic idea is to transform data in such a manner that formal guarantees, e.g. regarding the risk of singling out, linkage or inference, can be provided [2]. These formal guarantees are captured by so called *privacy models*. t -Closeness is a state-of-the-art model for protecting data from inference attacks. The model requires that the distribution of sensitive attribute values in a set of indistinguishable data records is not too different from the distribution of sensitive information in the overall dataset [3].

¹Corresponding Author: Raffael Bild, Institute of Medical Informatics, Statistics and Epidemiology, University Hospital rechts der Isar, Technical University of Munich, Ismaninger Str. 22, 81675 Munich, Germany; E-mail: raffael.bild@tum.de.

2. Objective

The ARX Data Anonymization Tool is among the few software solutions for quantitative data anonymization, that have found wide-spread adoption. ARX focuses on data transformation methods which have been specifically recommended for applications to health data and it implements models for protecting data from singling out, linkage and inference [4]. t -Closeness is amongst the models supported.

t -Closeness has been specified in different variants that apply to variables with different scales of measure. One of these variants focuses on variables which are totally ordered. This model is particularly relevant in practice, as it is one of the few privacy models which have been proposed for protecting sensitive numeric variables.

When using ARX to protect complex datasets using the t -closeness model, however, we realized that the initial implementation is not scalable. Upon further inspection, we realized that directly implementing a scalable algorithmic representation of the mathematical definition of the model t -closeness proves difficult in general. In this paper we therefore present a series of optimizations that we have developed to achieve efficiency in productive use. All of them have been integrated into ARX.

3. Methods

3.1. Problem Definition

t -Closeness is a condition that applies to equivalence classes, i.e. groups of records which are indistinguishable regarding attributes that could be used for linking records. Let $P(e) = (p_1, p_2, \dots, p_m)$ be the relative frequency distribution of sensitive values in a given equivalence class e and let $Q = (q_1, q_2, \dots, q_m)$ be the relative frequency distribution of sensitive values in the whole dataset. $D[P(e), Q]$ is the distance between the distributions $P(e)$ and Q [3]. It is defined as follows [3]:

$$D[P(e), Q] = \frac{1}{m-1} \sum_{i=1}^m \left| \sum_{j=1}^i (p_j - q_j) \right|.$$

A dataset fulfills t -closeness with numerical ground distance if for all equivalence classes e , $D[P(e), Q] \leq t$ holds.

	Age	Sex	LoS	AdmQtrr	Charge
Class e1	[25, 50[Male	[1, 5[3	50.000
	[25, 50[Male	[1, 5[3	60.000
Class e2	[50, 75[Female	[5, 10[1	60.000
	[50, 75[Female	[5, 10[1	60.000
	[50, 75[Female	[5, 10[1	70.000

Figure 1. Example discharge dataset. "LoS" = Length of stay, "AdmQtrr" = Admission quarter.

Figure 1 shows an example dataset with a sensitive attribute "Charge" and two equivalence classes $e1$ and $e2$ defined by the values of the other attributes. The distribution of sensitive values is $Q = (\frac{1}{5}, \frac{3}{5}, \frac{1}{5})$ since the values 50.000, 60.000 and 70.000 appear 1, 3 and 1 times in the whole dataset, respectively. For the equivalence classes, we get $P(e1) = (\frac{1}{2}, \frac{1}{2}, 0)$ and $P(e2) = (0, \frac{2}{3}, \frac{1}{3})$. Consequently, we have

$$D[P(e1), Q] = \frac{1}{2}(|\frac{1}{2} - \frac{1}{5}| + |\frac{1}{2} - \frac{1}{5} + \frac{1}{2} - \frac{3}{5}| + |\frac{1}{2} - \frac{1}{5} + \frac{1}{2} - \frac{3}{5} - \frac{1}{5}|) = \frac{1}{4},$$

$$D[P(e2), Q] = \frac{1}{2}(|-\frac{1}{5}| + |-\frac{1}{5} + \frac{2}{3} - \frac{3}{5}| + |-\frac{1}{5} + \frac{2}{3} - \frac{3}{5} + \frac{1}{3} - \frac{1}{5}|) = \frac{1}{6}.$$

Hence, we can conclude that the dataset satisfies t -closeness with $t = \max\{\frac{1}{4}, \frac{1}{6}\} = 0.25$.

A straight-forward implementation of t -closeness for fully ordered attributes would implement this by checking if the following inequality holds:

$$|r_1| + |r_1 + r_2| + |r_1 + r_2 + r_3| + \dots + |r_1 + \dots + r_{m-1}| \leq t(m-1)$$

Each r_i has the form $r_i = p_i - q_i$, where p_i is the frequency of the attribute value number i in the currently considered equivalence class e of the transformed data set, and q_i the frequency of the attribute value number i in the entire input dataset.

As we will show in Section 4 this process is highly inefficient. The main reason is that it needs to iterate over all sensitive attribute values contained in the overall dataset. Given that this process needs to be executed for all equivalence classes, the worst-case complexity is $O(n^2)$ where n is the number of records in the dataset.

3.2. Optimization Approaches

In this section, we present three optimizations that we used to improve our initial, straight-forward implementation of the model.

Optimization 1 – Fibonacci hashing: The first optimization addresses the implementation level. One of the most time-consuming aspects of implementing a check for t -closeness is to dynamically group the sensitive attribute values in each class to determine their frequency. The standard data structure used for this purpose are hash tables. ARX already used an efficient implementation provided by the High Performance Primitive Collections for Java library [5]. However, these collections are still much more complex than required, as they for example support updating the data stored in a map. We therefore implemented a simplified hash table using Fibonacci hashing based on the golden ratio to reduce the number of CPU cycles required for adding and querying elements.

Optimization 2 – Check pruning: The second optimization addresses the mathematical definition of the model. Let us assume that the attribute values number $1 \dots k$ in the equivalence class currently under consideration do not occur at all, then $p_1 = p_2 = \dots = p_k = 0$ holds and the first k summands in the above condition have the form:

$$|-q_1| + |-q_1 - q_2| + \dots + |-q_1 - \dots - q_k|.$$

This partial sum depends only on the input dataset and can therefore be pre-calculated for every possible value of k before the individual equivalence classes of the transformed dataset are checked.

These precalculations can be performed in an initialization step, for ascending values of k until the corresponding subtotal is greater than the threshold, i.e. the following holds:

$$|-q_1| + |-q_1 - q_2| + \dots + |-q_1 - \dots - q_k| > t(m-1).$$

Let us denote the smallest value of k for which this inequality is fulfilled with x .

When checking whether t -closeness holds for a specific equivalence class, as a first step, we calculate the smallest index of any attribute value occurring in the class. We call this index y . When $y > x$ it can be inferred that the following summands are included in the relevant sum:

$$|-q_1| + |-q_1 - q_2| + \dots + |-q_1 - \dots - q_x|.$$

It follows that the threshold $t(m-1)$ will definitely be exceeded and computations can already be stopped at this point (concluding that privacy guarantees are not fulfilled).

Optimization 3 – Summand pruning: The third optimization adds an additional pruning mechanism using pre-computations.

It can be used in cases where the pruning strategy described above is not applicable, i.e. if $y \leq x$ holds. It works by starting the summation at position y , using an appropriate sum which has been pre-calculated ahead of time for all $k \leq x$ as a starting point:

$$|-q_1| + |-q_1 - q_2| + \dots + |-q_1 - \dots - q_k|.$$

3.3. Experimental Design

To evaluate our approach, we used the following data from registries and health surveys from the U.S.: 100,937 records about traffic accidents from the NHTSA Fatality Analysis Reporting System (FARS), 539,253 records from the American Time Use Survey (ATUS) and 1,193,504 records from the Integrated Health Interview Series (IHIS). Moreover, we analyzed a subset of a synthetic discharge dataset which is particularly hard to protect from sensitive attribute disclosure (SPD) [6]. We also included two de-facto standard datasets for benchmarking anonymization methods: 30,162 records from the 1994 U.S. Census (ADULT) and 63,441 records from the 1998 KDD competition (CUP). For a detailed specification of the datasets we refer to [7].

We anonymized the datasets with attribute generalization and record suppression to produce output datasets which fulfill t -closeness for fully ordered attributes. We varied the risk threshold t ($0.5 \geq t \geq 0.1$) to study the effect of our optimizations on different parameterizations. As a baseline, we used our original, unoptimized implementation. In the software, both pruning strategies are combined into a common implementation and we therefore present one measurement capturing both of them.

4. Results

Figure 2 shows the results of our experiments. As can be seen, our optimizations improved execution times by up to a factor of more than two. Each optimization had a positive effect in all setups while the degree of effectiveness of each optimization varied between setups. Pruning was possible in between 69% and 99% of the checks performed, but the effect on execution times varied.

The differences in the impact of optimizations on execution times can be explained by considering the distribution of sensitive attributes values in the different datasets. The impact was lower for datasets with a small number of distinct values, i.e. ADULT (7), ATUS (7) and IHIS (10), higher for datasets with more distinct sensitive values, i.e. CUP (81) and FARS (20). We measured the strongest effect for the SPD dataset, which has 101 different sensitive attribute values. When the number of sensitive attribute values is high, execution times are also higher, implying the our optimizations are more effective exactly in the cases where they are needed the most.

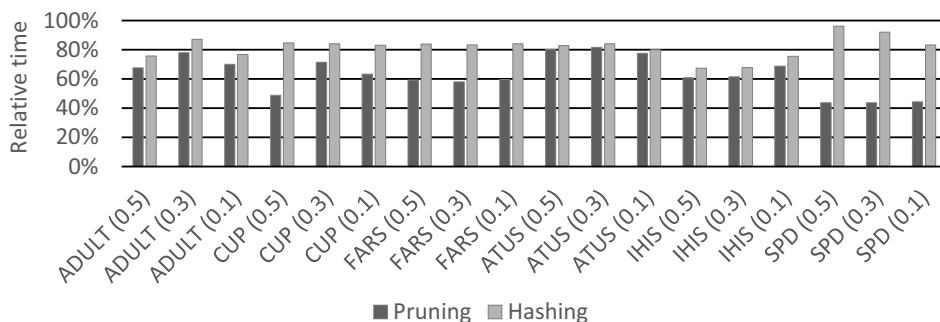


Figure 2. Execution times relative to the original implementation for various datasets and risk thresholds.

5. Conclusion

Due to its usability, flexibility and scalability, ARX is actively used in many areas, including commercial big data analytics platforms, medical research projects, clinical trial data sharing and for training purposes. An important reason for ARX's scalability are the many optimizations that have been integrated into the software. In prior work we have, for example, presented methods to improve the scalability of optimization algorithms for trading off risks vs. utility [8] and a versatile optimized runtime environment for anonymization algorithms [9].

In this paper, we have presented an optimization affecting a specific and important privacy model only. Our approach addresses the implementation level as well as the mathematical definition of the model implemented. Our solution utilizes high-performance data structures as well as pre-computation techniques. The model is particularly relevant in practice, as it is one of the few approaches which can be used to protect sensitive numeric data.

References

- [1] V. Gligorijević et al., Integrative methods for analyzing big data in precision medicine, *Proteomics* **16** (2016), 741–758.
- [2] B. C. M. Fung, K. Wang, A. W.-C. Fu and P. S. Yu, *Introduction to privacy-preserving data publishing: Concepts and techniques*, CRC Press, 1st ed., 2010, ISBN 9781420091489.
- [3] N. Li, T. Li and S. Venkatasubramanian, t-closeness: Privacy beyond k-anonymity and l-diversity, *2007 IEEE 23rd International Conference on Data Engineering*, IEEE, 2007, 106–115.
- [4] F. Prasser and F. Kohlmayer, Putting statistical disclosure control into practice: The ARX data anonymization tool, *Medical data privacy handbook*, Springer, 2015, 111–148.
- [5] S. Osinski and D. Weiss, HPPC: High Performance Primitive Collections for Java, <https://labs.carrotsearch.com/hppc.html>, accessed: October 10, 2019.
- [6] D. Sánchez, S. Martínez and J. Domingo-Ferrer, Comment on “Unique in the shopping mall: On the reidentifiability of credit card metadata”, *Science* **351** (2016), 1274–1274.
- [7] F. Prasser, F. Kohlmayer and K. A. Kuhn, A benchmark of globally-optimal anonymization methods for biomedical data, *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, IEEE, 2014, 66–71.
- [8] F. Prasser, F. Kohlmayer and K. A. Kuhn, Efficient and effective pruning strategies for health data de-identification, *BMC medical informatics and decision making* **16** (2016), 49.
- [9] F. Kohlmayer, F. Prasser, C. Eckert, A. Kemper and K. A. Kuhn, Highly efficient optimal k-anonymity for biomedical datasets, *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2012, 1–6.

Emerging Concepts and Applied Machine Learning Research in Patients with Drug-Induced Repolarization Disorders

Mina BJELOGRLIC^{a,1}, Arnaud ROBERT^a, Arnaud MIRIBEL, Mehdi NAMDAR^b,
Baris GENCER^{b,c}, Christian LOVIS^a and François GIRARDIN^d

^a*Division of Medical Information Sciences, University Hospitals of Geneva and
University of Geneva, Geneva, Switzerland*

^b*Cardiology Division, Department of Medicine, University Hospitals of Geneva and
University of Geneva, Geneva, Switzerland*

^c*Brigham and Women's Hospitals, TIMI study group, Harvard Medical School, Boston,
United States of America*

^d*Division of Clinical Pharmacology and Toxicology; Medical Direction, University
Hospitals of Geneva and University of Geneva, Geneva, Switzerland*

Abstract. The paper presents a review of current research to develop predictive models for automated detection of drug-induced repolarization disorders and shows a feasibility study for developing machine learning tools trained on massive multimodal datasets of narrative, textual and electrocardiographic records. The goal is to reduce drug-induced long QT and associated complications (Torsades-de-Pointes, sudden cardiac death), by identifying prescription patterns with pro-arrhythmic propensity using a validated electronic application for the detection of adverse drug events with data mining and natural language processing; and to compute individual-based predictive scores in order to further identify clinical conditions, concomitant diseases, or other variables that correlate with higher risk of pro-arrhythmic situations.

Keywords. Machine Learning, Clinical Decision Support System, Pharmacovigilance, Adverse Drug Events, Electrocardiography, Long QT, Torsades-de-Pointes, Repolarization Disorders, Analytic-Decision Modelling.

1. Introduction

Drug-induced ventricular arrhythmia is a major cause of iatrogenic morbidity and mortality [1], particularly in circumstances, where a wide range of psychotropic and illicit drugs are consumed (e.g. people who inject drugs, living in detention or with mental illness). Electrocardiographic (ECG) screening is a widely available method for detecting repolarization abnormalities. Long QT (LQT), defined as a QT interval duration of >470 ms for men and >480 ms for women, is the main surrogate marker for Torsades-de-Pointes (TdP), a pathognomonic, drug-induced polymorphic ventricular arrhythmia [2]. The probability to develop a TdP increases with the extent of the

¹Mina Bjelogrljic, PhD, Division of Medical Information Sciences, University Hospitals of Geneva and University of Geneva, Geneva, Switzerland. E-mail address: mina.bjelogrljic@hcuge.ch.

corrected QT interval (QTc) prolongation. Most TdP are transient and unsustained, but can degenerate to malignant ventricular fibrillation (VF) and sudden cardiac death (SCD).

Machine Learning (ML) based techniques are expected to support clinicians by allowing more accurate detection of abnormalities that would be missed using traditional diagnostic criteria (QT duration, axis, visual T-wave morphology). In the past decade, algorithms for pattern recognition and classification using learning techniques have become interesting in health applications for two main reasons; firstly, the digital information stored in Electronic Health Records (EHRs) is massively increasing [3], and there are available open databases, such as the MIT-BIH Arrhythmia Database [4]; secondly, the performance of ML algorithms have reached a point where either the task is better achieved by automatic systems, or further hidden underlying physiological phenomena can be re-addressed from a new analytical perspective. Recent research groups have tackled the problem of ECG-based heartbeat classification in order to detect arrhythmias [5]. ECG-based ML approaches include several steps after the acquisition of the ECG signals; preprocessing, segmentation, feature extraction, selection, and classification with learning algorithms [6].

2. Scope

Virtually all medications inducing LQT act by blocking the outward IKr current, which is mediated by the potassium channel encoded by the KCNH2 gene [7]. The measured QT value is further corrected according to a calculation based on heart rate (most frequently Bazett and Fridericia formulas), i.e. the QTc. Long QTc interval is a sensitive marker for TdP, but it is not specific since drug-induced QTc. More importantly, the lack of LQT specificity causes the discontinuation of valuable patients' treatment and new drug development [8]. Separate analysis of J-Tpeak and Tpeak-Tend intervals helps to identify medications which prolong the QTc interval with counterbalancing inward current block and limited risk of arrhythmias. Automated measurement methodology and algorithm for the evaluation of the early repolarization J-Tpeak and late repolarization Tpeak-Tend intervals were proposed to assess drug-induced proarrhythmia propensity, using a two-step T-wave delineation method [9]. A wide range of illicit substances and medications are associated with repolarization abnormalities. In psychiatry, risk factors for developing drug-induced TdP include chronic hepatitis C infection, HIV, electrolyte abnormalities, renal or hepatic dysfunction, pre-existing coronary heart disease, treatment with more than one QT-prolonging drug, older age, female sex, and genetic predisposition [10,11]. Abnormal electrocardiographic (ECG) recordings are reported in 27.3% of psychiatric inpatients at hospital admission and repolarization disorders were the most prevalent abnormal findings (11.8%).

3. Methodology

In the University Hospitals of Geneva (HUG), since 2017, all ECG devices (n=220) have been replaced by the same electrocardiographic devices (Philipps TC-70) and all records are automatically converted into a DICOM® [12] file (>8'000 ECGs per month performed in all care settings at HUG) and stored in the Picture Archiving Communication System (PACS) [13]. Computerized ECG health records with structured metadata in XML format are available to be extracted from the hospital DPI (Dossier

Patient Intégré) [14] secured servers for research purposes. All records can be analyzed with socio-demographic, clinical, laboratory, genetic, and drug prescription data at the individual patient level, using an agreed protocol and database structure. The latter cardiac information combined with electronic prescription systems and natural language processing (NLP), allow ECG feature extraction and drug imputability assessment, based on interactions, dose variations, and chronological relationship with drug exposition. Moreover, it stores all measurements from the recordings, meaning it includes boundary detection, epoch extraction, and characteristic measurements for each lead. In the proposed methodology (see **Figure 1**), electrographic, structured, and unstructured text data is collected and stored in a multimodal database, which will first pass a quality control, and then be used to extract and select relevant features to build models for RD prediction.

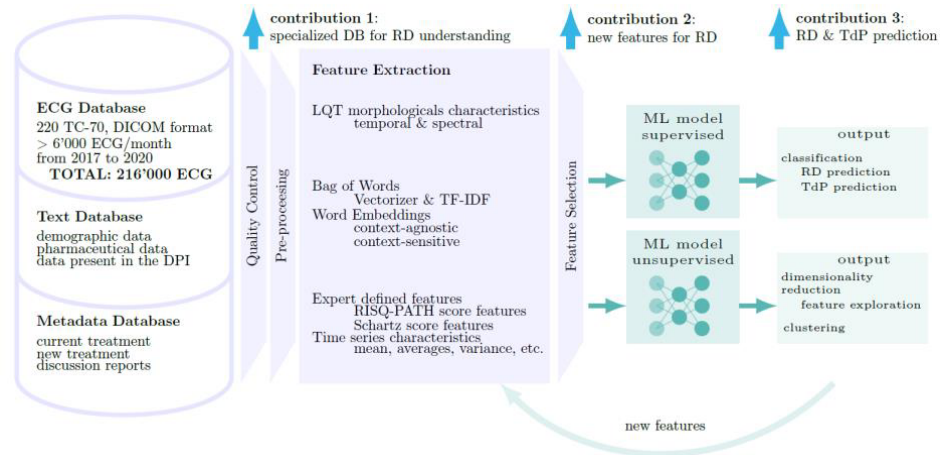


Figure 1. Research Methodology. Database (DB), Repolarization Disorders (RD), LQT (Long QT), Torsades de Pointe (TdP), Machine Learning (ML).

A feature set proposed in the literature include ECG-intervals, morphological (e.g. amplitude), Vector Cardiographs (VCG) [15], Temporal VCG (TVCG) [16], different transforms, including temporal polynomial fitting with Hermite Basis decompositions [17], and different versions of time-frequency representations, such as Discrete Wavelet Transforms (DWT) [18]. The classification, according to the resulting feature vectors is done using classifiers such as linear discriminant (LD), support vector machines (SVM) [19], Random Forests [20], and ensemble classifiers with the advantage of bagging both temporal and morphological feature-based classifiers [21]. Long-short-term memory (LSTM) networks, combined with a wavelet sequence (WS) [22], and algorithms with deep learning (DL) architectures [23] have been investigated. For inter-patient classification convolutional neural networks (CNN) and recurrent neural networks (RNN) give very promising results [24].

4. Feasibility study

There are no a priori exclusion criteria. Nonetheless, ECG of insufficient quality (not present QT measurements) or with inverted electrodes will not be included. With a

minimum 70'000 ECG recorded and saved in the hospital's database (PACS), the estimated amount of available ECG over the target period (three years) is 200'000 EC. **Figure 2** shows estimations, over a four week period, of QT density distributions depending on age and sex.

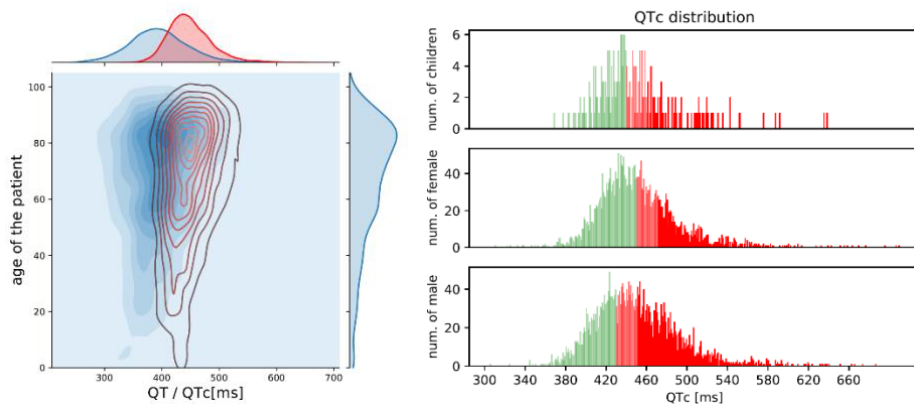


Figure 2. Left: density distributions estimation (based on a four-week period in 2019) of patient age and QT/QTc (Bazett's formula, in red) values of patients at HUG. Right: QTc (Bazett's formula) distributions over a four-week period in 2019 at HUG. For children (under the age of 15 years), women (above 15 years old, female sex), and men (above 15 years old, male sex). From green to dark red; normal, intermediate, and long QTc [ms].

5. Discussion

Approximately 30% of the inpatients are taking QT prolonging drugs, out of which 50% will have additional QT prolonging drugs [25]. In our preliminary analysis of ECG structured data (early 2019), RD prevalence and QT interval variability increased with patient age and the number of medications. The proposed research will exploit large databases to answer the central question of the influence of drugs on RD, particularly LQT and related severe complications. This question is of major importance since QT prolongation may lead to iatrogenic life-threatening arrhythmia, whose indicators could be detected to prevent consequences by instant electronic alert systems. ML-based approach could bring about innovation opportunities and new dimensions to comprehend iatrogenicity. Applied ML will influence decision-making processes (at individual and more global levels, such as health authorities), education, and training (including general practitioners, cardiologists, and medicine students). Additionally, CDSS can potentially decrease the number of misdiagnosed patients, the extra costs of non-suitable hospitalizations (40% of patients oriented to specialists for LQTS are misdiagnosed) [26], and prevent patients from severe cardiac events, such as arrhythmias or, possibly, sudden cardiac death.

References

- [1] Schwartz PJ, Woosley RL. Predicting the Unpredictable. *J Am Coll Cardiol* 2016 Apr;67(13):1639–1650. [doi: 10.1016/j.jacc.2015.12.063]

- [2] Beach SR, Celano CM, Sugrue AM, Adams C, Ackerman MJ, Noseworthy PA, Huffman JC. QT Prolongation, Torsades de Pointes, and Psychotropic Medications: A 5-Year Update. *Psychosomatics* 2018 Mar;59(2):105–122. [doi: 10.1016/j.psych.2017.10.009]
- [3] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE J Biomed Health Inform* 2018 Sep;22(5):1589–1604. [doi: 10/gddkw8]
- [4] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng Med Biol Mag* 2001 Jun;20(3):45–50. [doi: 10/c5zzg6]
- [5] Luz EJ da S, Schwartz WR, Cámara-Chávez G, Menotti D. ECG-based heartbeat classification for arrhythmia detection: A survey. *Comput Methods Programs Biomed* 2016 Apr;127:144–164. [doi: 10/f8fq6x]
- [6] Llamado M, Martínez JP. Heartbeat Classification Using Feature Selection Driven by Database Generalization Criteria. *IEEE Trans Biomed Eng* 2011 Mar;58(3):616–625. [doi: 10/d32zzt]
- [7] Itoh H, Crotti L, Aiba T, Spazzolini C, Denjoy I, Fressart V, Hayashi K, Nakajima T, Ohno S, Makiyama T, Wu J, Hasegawa K, Mastantuono E, Dagradi F, Pedrazzini M, Yamagishi M, Berthet M, Murakami Y, Shimizu W, Guicheney P, Schwartz PJ, Horie M. The genetics underlying acquired long QT syndrome: impact for genetic screening. *Eur Heart J* 2016 May 7;37(18):1456–1464. [doi: 10.1093/eurheartj/ehv695]
- [8] Haverkamp W. The potential for QT prolongation and proarrhythmia by non-antiarrhythmic drugs: clinical and regulatory implications. Report on a Policy Conference of the European Society of Cardiology. *Eur Heart J* 2000 Aug 1;21(15):1216–1231. [doi: 10.1053/ehj.2000.2249]
- [9] Johannesen L, Vicente J, Hosseini M, Strauss DG. Automated Algorithm for J-Tpeak and Tpeak-Tend Assessment of Drug-Induced Proarrhythmia Risk. Rasmussen RL, editor. *PLOS ONE* 2016 Dec 30;11(12):e0166925. [doi: 10/f9mn7f]
- [10] Girardin F, Sztajzel J. Cardiac adverse reactions associated with psychotropic drugs. *Dialogues Clin Neurosci* 2007 Mar;9(1):92–95.
- [11] Girardin F, Gex-Fabry M, Berney P, Shah D, Gaspoz J-M, Dayer P. Drug-Induced Long QT in Adult Psychiatric Inpatients: The 5-Year Cross-Sectional ECG Screening Outcome in Psychiatry Study. *Am J Psychiatry* 2013 Dec;170(12):1468–1476. [doi: 10.1176/appi.ajp.2013.12060860]
- [12] DICOM specification website [Internet]. Available from: <https://www.dicomstandard.org/>
- [13] Rosset C, Rosset A, Ratib O. General Consumer Communication Tools for Improved Image Management and Communication in Medicine. *J Digit Imaging* 2005 Dec;18(4):270–279. [doi: 10.1007/s10278-005-6703-2]
- [14] Lovis C, Baud RH, Planche P. Power of expression in the electronic patient record: structured data or narrative text? *Int J Med Inf* 2000;58–59:101–110. [doi: 10/fnmvqb]
- [15] Soria ML, Martínez J. Analysis of multidomain features for ECG classification. 2009 36th Annu Comput Cardiol Conf CinC 2009. p. 561–564.
- [16] Garcia G, Moreira G, Menotti D, Luz E. Inter-Patient ECG Heartbeat Classification with Temporal VCG Optimized by PSO. *Sci Rep* 2017 Sep 5;7(1):10543. [doi: 10/gbwznc]
- [17] Li BN, Dong MC, Vai MI. Modelling cardiovascular physiological signals using adaptive Hermite and wavelet basis functions. *IET Signal Process* 2010;4(5):588. [doi: 10/bmr8fv]
- [18] Daamouche A, Hamami L, Alajlan N, Melgani F. A wavelet optimization approach for ECG signal classification. *Biomed Signal Process Control* 2012 Jul;7(4):342–349. [doi: 10/dmp2db]
- [19] Kanaan L, Merheb D, Kallas M, Francis C, Amoud H, Honeine P. PCA and KPCA of ECG signals with binary SVM classification. 2011 IEEE Workshop Signal Process Syst SIPS [Internet] Beirut, Lebanon: IEEE; 2011 [cited 2019 Mar 4]. p. 344–348. [doi: 10/fdppnv]
- [20] Li T, Zhou M. ECG Classification Using Wavelet Packet Entropy and Random Forests. *Entropy* 2016 Aug 5;18(8):285. [doi: 10/f822bk]
- [21] Mondéjar-Guerra V, Novo J, Rouco J, Penedo MG, Ortega M. Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. *Biomed Signal Process Control* 2019 Jan;47:41–48. [doi: 10/gfxjrr]
- [22] Yildirim Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput Biol Med* 2018 May;96:189–202. [doi: 10/gc8srm]
- [23] Pyakillya B, Kazachenko N, Mikhailovsky N. Deep Learning for ECG Classification. *J Phys Conf Ser* 2017 Oct;913:012004. [doi: 10/gfxjrk]
- [24] Guo L, Sim G, Matuszewski B. Inter-Patient ECG Classification with Convolutional and Recurrent Neural Networks. *ArXiv181004121 Eess Q-Bio* 2018 Sep 27;11.
- [25] Zürcher DJ-P, Schlaepfer J, Waeber G, Pasquier M. Le QT long acquis. *Rev Médicale Suisse* 2013;9:1538–42.
- [26] Taggart NW, Haglund CM, Tester DJ, Ackerman MJ. Diagnostic miscues in congenital long-QT syndrome. *Circulation* 2007 May 22;115(20):2613–2620. PMID:17502575

End-to-End Approach for Structuring Radiology Reports

Kento SUGIMOTO^{a,1}, Toshihiro TAKEDA^a, Shoya WADA^a, Asuka YAMAHATA^a,
Shozo KONISHI^a, Shiro MANABE^a and Yasushi MATSUMURA^a

^a *Department of Medical Informatics, Osaka University Graduate School of Medicine*

Abstract. Radiology reports include various types of clinical information that are used for patient care. Reports are also expected to have secondary uses (e.g., clinical research and the development of decision support systems). For secondary use, it is necessary to extract information from the report and organize it in a structured format. Our goal is to build an application to transform radiology reports written in a free-text form into a structured format. To this end, we propose an end-to-end method that consists of three elements. First, we built a neural network model to extract clinical information from the reports. We experimented on a dataset of chest X-ray reports. Second, we transformed the extracted information into a structured format. Finally, we built a tool that enabled the transformation of terms in reports to standard forms. Through our end-to-end method, we could obtain a structured radiology dataset that was easy to access for secondary use.

Keywords. Natural Language Processing, Radiology Report, Information Extraction

1. Introduction

Radiology reports are created by radiologists to communicate with referring clinicians, and play an important role in patient care. Radiology reports are usually written in free-text format. It has been reported that the ambiguity of terminology or style in free-text can reduce the clarity of the report, causing inaccurate communication [1]. European Society of Radiology has taken the initiative in developing structured reporting to improve the quality of radiology reports [2]. They mention structured reporting also has the potential to facilitate clinical research and the development of radiological applications.

There are a lot of studies on information extraction from radiology reports [3]. Some studies examined Japanese radiology reports [4,5]. In recent years, extraction method using a deep learning has drawn much attention [6]. However, most studies have only focused on the extraction process and did not refer to the structuring process.

In this study, we propose an end-to-end approach for structuring radiology reports. Our aim is utilizing massive amounts of unstructured reports as training data for a lesion detection system [7]. Specifically, we take three steps: *extraction*, *structuring* and *normalization*. In the first step, we build a recurrent neural network-based model for entity recognition. In the second step, we organize the extracted information for storage in a database. The data is stored in a tabular database for accessibility. It is well-known that

¹ Corresponding Author, Osaka University, 1-1 Yamadaoka Suita, Osaka, Japan;
E-mail: sugimoto.kento@hp-info.med.osaka-u.ac.jp

radiology reports contain heterogeneous writing styles, including non-standard terminology and abbreviations [8]. In the third step, we transform non-standard terms in the report to a standard form.

2. Material and methods

2.1. Radiology dataset

In this study, we used chest X-ray reports from 2000 to 2017 that were stored in the radiology information system at Osaka University Hospital. The dataset consisted of 319,130 reports, most of which were written in Japanese. This study was approved by the institutional review board of Osaka University Hospital (Permission number: 17166).

2.2. Entity recognition

2.2.1. Gold Standard

We randomly sampled 5,000 reports for a gold standard dataset. A report was segmented into a single sentence and word tokenization was implemented using MeCab [9].

One non-medical expert was responsible for the annotation process. Two medical experts (a clinician and a radiological technologist) reviewed the results.

A previous study defined 5 semantic classes (*Anatomy*, *Anatomy Modifier*, *Observation*, *Observation Modifier*, and *Uncertainty*) [10]. Another defined 4 semantic classes (*Clinical Finding*, *Body Location*, *Descriptor* and *Medical Device*) [6]. With reference to previous studies, we defined 3 semantic classes: *Clinical Finding* (CF), *Body Location* (BL), and *Body Location Modifier* (BLM). CF includes terms related to observation and abnormalities in the report. BL includes terms related to the anatomical area. BLM encompasses terms that modify the BL. To differentiate the certainty of the Clinical Findings, they were split into three labels: *Clinical Finding Positive* (CFP), *Clinical Finding Suspicious* (CFS), *Clinical Finding Negative* (CFN). Figure 1 shows an example of annotation.

A part-solid nodule is seen in the left upper lung fields.
CFP BLM BL

Figure 1. annotation example

2.2.2. Neural Network Architecture

Our deep learning model is based on the architecture described by Ma and Hovy [11]. The encoder part of this model is composed of the Bidirectional Long Short Term Memory (BiLSTM) [12] layers. The character representation of each word computed by convolutional neural networks (CNNs) is concatenated with the word representation. Finally, the output of BiLSTM is fed to the CRF [13] layers to decode the label sequence.

2.3. Structuring

Entities from the report need to be transformed to tabular format records. We created simple rules using position information of the entities. In the common word order of radiology reports written in Japanese, BL appears after BLM, before CF. Based on this assumption, pairs of BLM and BL, BL and CF were created (Figure 2).

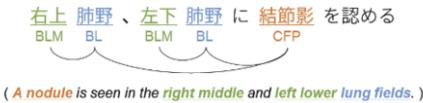


Figure 2. an example of pairs in the report

When transforming, a pair of BLM term and BL term was concatenated as one term, and a CF was divided into CF and Certainty. Multiple records were created from one sentence when BL and CF had a one-to-many (many-to-one) relationship (Figure 3).

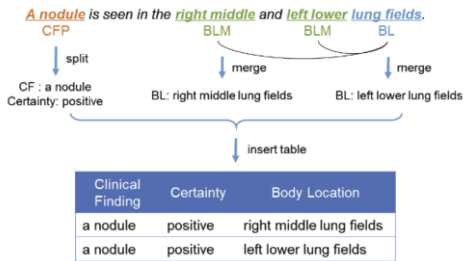


Figure 3. an example of creating records from the extracted entities.

2.4. Normalization

Each term needs to be normalized automatically. We created a concept table from scratch since there is no comprehensive radiology vocabulary in Japan. This table has 2 columns; *mention* and *concept*. Values of mention column are terms that occurred in the radiology reports, and they were collected from the outputs of our entity extraction model (As we described in 2.2.2). Value of concept column is the standard form of its mention column.

3. Results

3.1. Entity recognition

Five thousand reports were divided into 4,000 for training, 500 for development, and 500 for validation. Hyper-parameters were tuned by using a development dataset. We used

pre-trained word vectors to help the learning process. Pre-trained word vectors were obtained using Contiguous Bag of Words (CBOW) [14] from 317,130 unlabeled radiology reports. We used the RMSProp algorithm [15] with a batch size of 32. A 0.5 dropout rate was applied to avoid overfitting.

Our experimental results are shown in Table 1. Our model achieves 0.938 in the F1-score, which shows that the model can accurately extract target entities.

Table 1. Performance metrics of each semantic class

Semantic classes	Precision	Recall	F1-score	No. of entities
Body Location Modifier	0.953	0.968	0.960	4,024
Body Location	0.952	0.959	0.955	7,205
Clinical Finding Positive	0.904	0.920	0.912	7,574
Clinical Finding Suspicious	0.874	0.902	0.888	1,672
Clinical Finding Negative	0.992	0.932	0.961	2,818
Total	0.937	0.941	0.938	23,297

3.2. Structuring

We manually created 453 records from 200 reports. We evaluated the accuracy of the structuring process assuming that all entities were correctly extracted. Our structuring processor was able to correctly transform 436 records (Accuracy is 0.96).

3.3. Normalization

We created a concept table for normalization. This table was created by some medical experts (clinicians and radiological technologists). We created 21 concept terms for BL by dividing chest X-ray images. And, the concept for CF was decided by picking up frequently occurring words that were extracted as CF entity. We then combined similar concepts into one concept term. Finally, we defined 121 concept terms.

4. Discussion

In Table 1, we show the performance metrics of each semantic class. The F1-score of semantic classes exceeded 0.9, excluding CFS. The lower F1-score in CFS was related to the number of entities in the training dataset. Table 1 shows the number of entities in the CFS group was lower in comparison to other semantic classes.

As we described in 3.2, the simple rule-based method is applicable for structuring reports. This result may be due to the high similarity of the writing style of the reports.

The present study was associated with some limitations. First, some important information could not be covered in our semantic classes. For example, the size and shape of the lesions were not included, even though they are valuable for clinical research. Second, we only used an in-house dataset, which means that the performance of our method might differ if a dataset from a different hospital was used. Third, although our concept terms of BL and CF were 21 and 121, respectively, the granularity is not general and may not be appropriate for some secondary uses.

5. Conclusion

In this paper, we built an end-to-end method for transforming unstructured radiology reports into a structured format. We combined machine learning for entity recognition with a rule-based method for structuring and normalization to pursue a better approach. While our method leaves room for improvement of versatility, the present study shows the way to facilitate secondary use of radiology reports.

Acknowledgments

This work was supported by Council for Science, Technology and Innovation (CSTI), cross-ministerial Strategic Innovation Promotion Program (SIP), "Innovative AI Hospital System" (Funding Agency: National Institute of Biomedical Innovation, Health and Nutrition (NIBIOHN)).

References

- [1] J.M.L. Bosmans, J.J. Weyler, P.M. Parizel. Structure and content of radiology reports, a quantitative and qualitative study in eight medical centers. *European Journal of Radiology*. **72** (2009),354–8.
- [2] European Society of Radiology (ESR). ESR paper on structured reporting in radiology. *Insights Imaging*. **9** (2018),1–7.
- [3] E. Pons, L.M.M. Braun, M.G.M. Hunink, J.A. Kors. Natural Language Processing in Radiology: A Systematic Review. *Radiology*. **279** (2016),329–43.
- [4] T. Imai, Y. Onogi. Extracting numeric measurements and temporal coordinates from Japanese radiological reports. *Proceedings of SPIE - The International Society for Optical Engineering*. (2004),268–76.
- [5] T. Imai, E. Aramaki, M. Kajino, K. Miyo, Y. Onogi, K. Ohe. Finding malignant findings from radiological reports using medical attributes and syntactic information. *Stud Health Technol Inform*. **129** (2007),540–4.
- [6] S. Cornegruta, R. Bakewell, S. Withey, G. Montana. Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks. *Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis (LOUHI)*. (2016),17–27.
- [7] C. Wang, A. Elazab, J. Wu, Q. Hu. Lung nodule classification using deep feature fusion in chest radiography. *Computerized Medical Imaging and Graphics*. **57** (2017),10–8.
- [8] B. Reiner, E. Siegel. Radiology reporting: Returning to our image-centric roots. *American Journal of Roentgenology*. **187** (2006),1151–5.
- [9] T. Kudo. MeCab. <http://taku910.github.io/mecab/>.
- [10] S. Hassanpour, C.P. Langlotz. Information extraction from multi-institutional radiology reports. *Artif Intell Med*. **66** (2016),29–39.
- [11] X. Ma, E. Hovy. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. (2016),1064–74.
- [12] A. Graves, J. Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*. **18** (2005),602–10.
- [13] J. Lafferty, A. McCallum, F.C.N. Pereira, F. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *Proceeding of ICML*. (2001).
- [14] T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space. *ICLR*. (2013).
- [15] Y.N. Dauphin, H. de Vries, Y. Bengio. Equilibrated adaptive learning rates for non-convex optimization. *Proceeding NIPS'15 Proceedings of the 28th International Conference on Neural Information Processing Systems*. (2015),1504–12.

Evaluation of Document Retrieval Systems on a Medical Corpus in French: Indexation vs. Feature Learning

Arnaud ROBERT^{a,1}, Francis DAMACHI^{a,1}, Mina BJELOGRLIC^{a,1}, Jean-Philippe GOLDMAN^{a,1}, and Christian LOVIS^{a,1}

^a*Division of Medical Information Sciences, University Hospitals of Geneva and University of Geneva, Geneva, Switzerland*

Abstract. This paper presents five document retrieval systems for a small (few thousands) and domain specific corpora (weekly peer-reviewed medical journals published in French) as well as an evaluation methodology to quantify the models performance. The proposed methodology does not rely on external annotations and therefore can be used as an ad hoc evaluation procedure for most document retrieval tasks. Statistical models and vector space models are empirically compared on a synthetic document retrieval task. For our dataset size and specificities the statistical approaches consistently performed better than its vector space counterparts.

Keywords. Medical Information Retrieval, Document Retrieval, Natural Language Processing, Word Embedding.

1. Introduction

The field of Information Retrieval (IR) is a well-studied and yet still critical field of research. In the last few decades, additional effort has been made in the biomedical domain with the digitalization of medical data and the multiplication of biomedical scientific literature [1,2]. The goal of the present paper is to give a clear methodology in evaluating several IR systems on small (few thousands) and domain specific corpora (medical literature in French). For a specific query, document ranking is determined by a numerical score assigned by the IR system [3]. On the one side, early IR systems in the early 60's were based on Boolean and statistical approaches [4]. In term-based approaches documents are represented by a predefined arbitrary feature with no encoded semantic information (words are represented by an index in a vocabulary). On the other side, vector space (VS) models represent documents as a composition of vector representation of terms [5]. In VS each word is encode as a vector in a high dimensional space allowing to encode semantic information. Independently of the representation

¹ Authors contributed equally to this work, Division of Medical Information Sciences, University Hospitals of Geneva and University of Geneva, Geneva, Switzerland E-mail: Firstname.Lastname@hcuge.ch

used, most Document Retrieval systems compute a ranking between a query and all the documents in the corpus [3]. The documents with the highest score (rank) are considered most relevant and are returned [6,7]. Statistical models (*Boolean* and *Weighted*), vector space models (*Word2Vec-Mean* and *Word2Vec-Weighted*, *Word2Vec-Max*), are compared on a dataset consisting of 4,201 articles of weekly peer-reviewed medical journals published in French. All models are evaluated empirically on the same document retrieval task.

2. Data/corpus

The source of the dataset is the *Revue Medical Suisse* (RMS). The dataset consists of 4,201 articles of weekly peer-reviewed medical journals published in French in XML format spanning from years 2014 to 2019. Sections of the document such as the abstract, title, body, conclusion and references can be distinguished by specific XML tags. In order to evaluate the models performance we kept articles for which title, abstract and body were available and in French, which resulted into 1,401 documents. All models benefited from the same preprocessing steps. The first step consists into the extraction of the documents from the XML format to plain text. Then, each document is converted into lemmatized tokens (49,573 unique terms). Lemmatization is done by *Spacy*'s pre-trained statistical models for French [8]. Stopwords are removed (they do not add any semantic information), verbs are removed (add small semantic information), and word that occurs less than 2 were removed (to filter out mostly misspellings and extremely rare terms). Following this procedure, approximately 42% of words were filtered out, leaving a vocabulary size of 28,791 terms.

3. Methodology

We investigate different methodologies to rank documents according to a query. The proposed models fall in one of the two following categories: Term-based or VS models. The term-based models evaluated (*Boolean* and *Weighted* models) represents words as an index in a reference vocabulary while VS models represent words as vector in a high dimensional space [5], which encode semantic properties of the word. The evaluated VS models (*Word2Vec-Mean*, *Word2Vec-Weighted*, *Word2Vec-Max*) represent each term as a m -dimensional vector v using the well-known *Word2Vec* algorithm [5], where m is a user chosen parameter (called embedding size).

3.1. Term-based Models

In the *Boolean model*, for a vocabulary size of n words, the model encodes a text as an n -dimensional vector containing a '1' for each word present in the text and '0' otherwise. In order to evaluate the similarity between a query q and a document d the model first normalizes the two vectors: $\bar{q} = \frac{q}{\sum_i q_i}$, $\bar{d} = \frac{d}{\sum_i d_i}$, and then compute the cosine similarity between \bar{q} and \bar{d} , where the cosine similarity between two vectors a and b is defined as follow:

$$\text{sim}(a, b) = \frac{\sum_i a_i b_i}{\sqrt{\sum_i a_i^2 \cdot \sum_i b_i^2}} \quad (1)$$

The main drawback of the *Boolean model* is that it gives the same importance to all the words in a document. To include the word importance, the *Weighted model* uses the TF-IDF (for term frequency - inverse document frequency) metric that weight the importance of a term in a document. More formally TF-IDF will compute the weight of a term as follow:

$$\begin{aligned} \tilde{w}_d(t) &= \text{freq}_d(t) * (\log \frac{D+1}{d(t)+1} + 1), \\ w_d(t) &= \frac{\tilde{w}_d(t)}{\sum_{t'} \tilde{w}_d(t')}, \text{ with } \tilde{w}_d(t) = \text{freq}_d(t) * (\log \frac{D+1}{d(t)+1} + 1) \end{aligned} \quad (2)$$

where $\text{freq}_d(t)$ denotes the number of times the term t appears in the document d , i.e the term frequency and $d(t)$ denotes the number of documents that contain the term t , i.e. the document frequency. Query and documents are compared using cosine similarity (see eq. 1) [9].

3.2. Vector Space Models

For all VS models word representation was computed with *Word2Vec* algorithm, with an embedding size of 32, a context window size of 7, and 64 negative samples.

In *Word2Vec-Mean model*, documents are represented as the mean of their word vectors:

$d_j = \frac{1}{|d_j|} \sum_{i \in d_j} v_i$, where v_i is the m -dimensional vector representing word i . Queries

are represented with the same methodology. For a given query, the relevance of a document is computed with the cosine similarity (see eq. 1). The main drawback of the previous method is that it assigns the same weight to all words of the document. An alternative (*Word2Vec-Weighted model*) is to use the TF-IDF weights (see eq. 2), and

represent a document as follow: $d_j = \frac{1}{\sum_i w_{d_j(i)}} \sum_{i \in d_j} w_{d_j(i)} v_i$

When representing a document with the above methods, the mean can involve a potentially large number of vector (several hundreds), hence drowning the information in the mass. The *Word2Vec-Max model* solves this problem by focusing on the few words that are the most relevant regarding the query. For each term of the query q the algorithm find the closest term (w.r.t. the cosine similarity, see eq. 1) in the document and average them to compute the final similarity score: $\sum_{q_i \in q} \max_{d_j \in d} (\text{sim}(q_i, d_j))$.

4. Experiments

The dataset for testing consists of 1400 medical articles from RMS Journal. In order to evaluate the above information retrieval systems we used an article's title as a query. The precision of the model is then evaluated, by looking if the corresponding article appears in the top-N most relevant documents (N=1, 5, 10). Performance of some methods might fluctuate depending on the length of the documents composing the corpus, therefore, experiments were conducted on 3 different corpora:

- Corpus A, contains only the abstracts of the articles (25 to 184 words);

- Corpus B, contains only the body of the articles (431 to 8,839 words);
- Corpus A+B, contains both the abstract and the body of the articles.

Titles, used as queries, have from 1 to 29 words (mean=9.7 words) and the total number of words is the whole corpus is 1,853,939 (in titles + abstracts + bodies).

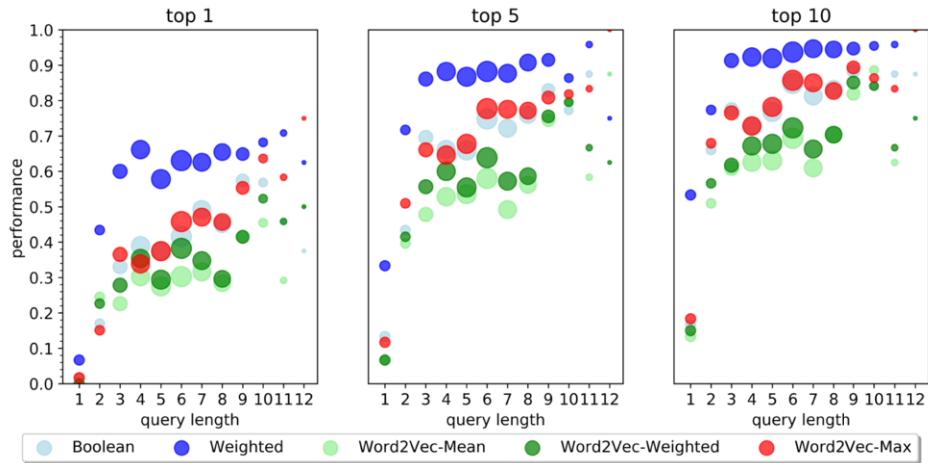


Figure 1. Performance according to query length. From left to right, it shows the top1, top5, and top10 performance. Models are color coded, and the size of the marker is proportional to the support size (i.e. number of queries).

For all the documents, all models, and the three different corpuses, the top1, top5, and top10 performances are shown in Table 1. The best performance for almost all sub-tasks is obtained by the *Weighted model*. In order to quantify the impact of the query length, Figure 1 shows the performance of each model according to the query length. The dominance of the *Weighted model* (in dark blue) is confirmed by this new perspective. The performance is the ratio between the number of queries where the correct document was in a specific ranking (i.e. top1, top5, top10) and the total number of queries. We achieve up to 94.5% of retrieval with top10 metric for 7 concept long queries.

Table 1. Shows the top1/top5/top10 performance of the five models on 1400 different queries

Method	Abstracts only	Bodies only	Abstracts + bodies
Boolean	801/1071/1138	564/963/1084	595/979/1092
Weighted	876/ 1186/1240	838 /1195/ 1276	861/1218/1295
Word2Vec- Mean	422/785/920	405/749/900	421/766/915
Word2Vec- Weighted	486/804/949	458/812/941	474/831/955
Word2Vec-Max	882 /1135/1192	572/982/1098	598/1004/1114

5. Discussion

First, it is important to note that the task is prone to advantage index-based methods since it is very likely that words that appear in the query (title) will occur in the target article. Hence, it might explain the very good results obtain by index based methods. Secondly

it seems that VS space models that compute document representation using all terms in the documents fail to provide a good representation, as suggested by the poor results obtained by *Word2Vec-Mean* and *Word2Vec-Weighted models*. This is also confirmed by the performance gain seen with *Word2Vec-Max* (which use less terms to represent documents).

6. Conclusion

This paper shows and evaluates how various document representations can impact the performance of a Document Retrieval system. The presented models are separated into two groups: term based and vector space models. In total we described and evaluated 5 models: *Boolean*, *Weighted*, *Word2Vec-Mean*, *Word2Vec-Weighted*, *Word2Vec-Max models*. The relevance of a document with respect to a query was computed using the cosine similarity between their appropriate representations. By varying the query length we see how the models perform in retrieving the top 1/5/10 documents.

Future work should focus on: (i) evaluating the presented methods against a much larger corpus (rather than only 1400 articles); (ii) investigating more in depth document representation (the *Word2Vec Max* use only a small subset of words, as many as the query length, to represent a document, this information loss impact the severely the document representation); (iii) efficient implementation of those systems to scale to larger database. The methodology proposed in this paper can be used as an ad hoc evaluation procedure for most document retrieval tasks in small and medium size text databases, in order to adopt the most effective IR strategy.

References

- [1] Roberts K, Simpson M, Demner-Fushman D, Voorhees E, Hersh W. State-of-the-art in biomedical literature retrieval for clinical cases: a survey of the TREC 2014 CDS track. *Inf Retr J* 2016 Apr 1;19(1):113–148. [doi: 10/ggbx8s]
- [2] Goeuriot L, Jones GJF, Kelly L, Müller H, Zobel J. Medical information retrieval: introduction to the special issue. *Inf Retr J* 2016 Apr;19(1–2):1–5. [doi: 10/ggbx9j]
- [3] Singhal A. Modern Information Retrieval: A Brief Overview. *Bull IEEE Comput Soc Tech Comm Data Eng* 2001;24:35–43.
- [4] Bordogna G, Pasi G. A fuzzy linguistic approach generalizing Boolean Information Retrieval: A model and its evaluation. *J Am Soc Inf Sci* 1993;44(2):70–82. [doi: 10/cvk4q4]
- [5] Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. *ArXiv13013781 Cs [Internet]* 2013 Jan 16 [cited 2019 May 17]; Available from: <http://arxiv.org/abs/1301.3781>
- [6] Kusner MJ, Sun Y, Kolkin NI, Weinberger KQ. From Word Embeddings To Document Distances. *Int Conf Mach Learn* 2015;957–966.
- [7] Ganguly D, Roy D, Mitra M, Jones GJF. Word Embedding based Generalized Language Model for Information Retrieval. *Proc 38th Int ACM SIGIR Conf Res Dev Inf Retr - SIGIR 15 [Internet]* Santiago, Chile: ACM Press; 2015 [cited 2019 Sep 23]. p. 795–798. [doi: 10/gf8tnc]
- [8] French · spaCy Models Documentation [Internet]. French. [cited 2019 Aug 7]. Available from: <https://spacy.io/models/fr>
- [9] Xu B, Lin H, Lin Y, Ma Y, Yang L, Wang J, Yang Z. Improve Biomedical Information Retrieval Using Modified Learning to Rank Methods. *IEEE/ACM Trans Comput Biol Bioinform* 2018 Nov 1;15(6):1797–1809. [doi: 10/ggbx8z]

Evaluation of the Quality of French Hospital Data for Perinatal Algorithms

Karine GOUESLARD ^a, Jonathan COTTENET ^a, Eric BENZENINE ^a, Pascale TUBERT-BITTER ^b, Catherine QUANTIN ^{a, b, c1}

^a*Biostatistics and Bioinformatics (DIM), University Hospital, Dijon, France;
Bourgogne Franche-Comté University, Dijon, France*

^b*Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases, Inserm, UVSQ, Institut Pasteur, Université Paris-Saclay, Villejuif, France
Inserm, CIC 1432, Dijon, France ; Dijon University Hospital, Clinical Investigation Center, clinical epidemiology/ clinical trials unit, Dijon, France*

Abstract. The aim of our validation study was to assess the quality of hospital data for perinatal algorithms on a national level. In each hospital, we selected 150 discharge abstracts of delivery (after 22 weeks of gestation), in 2014, and their corresponding medical records. Overall, 23 hospitals were included and 3,246 discharge abstracts were studied.

This first national validation study of several case-finding algorithms using various perinatal variables suggests that the French national hospital discharge abstracts database is an appropriate data source for epidemiological studies.

Keywords: hospital data, perinatal indicators, validation study.

1. Introduction

Administrative data, such as hospital or care consumption, that are collected primarily for reimbursement, contain several clinical diagnoses and procedures, which can be used to provide epidemiologic information (1). In France, 99.6% of the 800,000 annual births in France take place in hospitals (2). This renders the French hospital database particularly interesting for the investigation of perinatal. Nonetheless, the researchers using the medico-administrative databases are required to evaluate data quality via a validation study (3–7). In validation studies of case-finding algorithms, validity indices are used to assess data quality by comparing an algorithm in the simpler and easier-to-use data set versus a more elaborate, robust and reliable data set that is considered to be a “gold standard”. Furthermore, other validity indices are used to aid the selection of the best algorithms. In contrast to the less frequently conducted national surveys, the use of routinely collected health data in hospitals saves time and money when identifying infrequent and unfavorable delivery outcomes and improves health surveillance of women and their offspring (1).

The aim of our validation study was to assess the hospital data quality for perinatal algorithms on a national level.

¹ Corresponding Author, Catherine QUANTIN, CHU de Dijon - Service de Biostatistique et d'Informatique Médicale - BP 77908, 21079 Dijon CEDEX, France. catherine.quantin@chu-dijon.fr

2. Methods

The principle of this transversal multicenter study was to compare discharge abstract data entered into the French hospital database with the corresponding data from medical records, where we considered the latter to be the gold standard.

2.1. Population

Fifty health hospitals with a maternity unit were randomly selected in metropolitan France (other than Paris and Paris region), irrespective of the level of the unit. We developed a software program to randomly select 150 perinatal discharge abstracts per hospital (≥ 22 weeks of gestation) in 2014, from all discharge abstracts that contained a Z37 code and/or a delivery procedure. The same investigator collected data in each hospital from hospital paper-based or electronic medical records of pregnancy (prenatal care, the delivery and the post-delivery stay, and a discharge letter). Then, a comparison between data from medical records and data from discharge abstracts was made.

2.2 Statistical analysis

We explored several algorithms, including different combinations of codes in discharge abstracts from delivery stay, pregnancy stay, or hospitalization over the past 2 years before delivery. To evaluate the quality of the hospital database of discharge abstracts versus the hospital medical records (our gold standard), two validity indices the positive predictive value (PPV) and the sensitivity were calculated for dichotomous data. Continuous data were assessed by the concordance rate as the validity index. The rates of false negatives (FN) and false positives (FP) were also calculated in order to select the best algorithms with regards to the likelihood ratio (balancing specificity and sensitivity). This study was approved by the French National Committee for Data Protection (registration number 9132091). To meet the requirement of the data protection agency, the family names and first names were removed, the date of birth was replaced by the age at delivery, and the dates of admission and discharge were replaced by the length of stay.

3. Results

Twenty-two hospitals were finally included. Among them, 5 maternity units had less than 1,000 deliveries, 4 units had between 1,000 and 2,000 deliveries, 5 units had between 2,000 and 3,000 deliveries and 8 maternity units had more than 3,000 deliveries. Overall, 3,246 discharge abstracts were compared with their corresponding medical records.

3.1. Maternal indicators

The concordance rate of maternal age at delivery was 94.8%, with a 95% confidence interval (CI) [93.8-95.4]. The maternal characteristics and several types of morbidity are presented in Table 1.

Table 1: Quality of the hospital database of discharge abstracts versus the hospital medical records

	Medical records		Discharge abstracts		PPV		FP		FN		Sensitivity	
	n	%	n	%	%	95% CI	n	%	n	%	%	95% CI
Maternal characteristics and comorbidities												
Parity												
Primiparous women	981	40.2	971	39.8	93.3	92.3-94.3	65	2.7	75	3.1	92.4	91.3-93.5
Multiparous women	1,459	59.8	1,458	59.8	95.5	94.7-96.3	66	2.7	67	2.7	95.5	94.7-96.3
Overweight or obesity ≥ 25 kg/m ² *	1,104	37.6	220	7.5	98.6	98.2-99.0	3	0.1	887	30.2	19.7	18.3-21.1
Obesity ≥ 30 kg/m ² *	507	17.3	216	7.4	91.7	90.7-92.7	18	0.6	309	10.5	39.1	37.3-40.9
Uterus scar**	464	14.4	384	11.9	94.8	94.0-95.6	20	0.6	100	3.1	78.5	77.1-79.9
Diabetes Mellitus												
Type 1 diabetes ***	98	3.0	143	4.4	50.4	48.7-52.1	71	2.2	26	0.8	73.5	72.0-75.0
Type 2 diabetes ***	51	1.6	31	1.0	67.7	66.1-69.3	10	0.3	30	0.9	41.2	39.5-42.9
Type 1 or type 2 diabetes	148	4.6	170	5.2	65.9	64.3-67.5	58	1.8	36	1.1	75.7	74.2-77.2
High blood pressure	29	0.9	34	1.0	32.4	30.8-34.0	23	0.7	18	0.6	37.9	36.2-39.6
Pregnancy-related disorders												
Gestational diabetes												
Code O24.4	482	14.9	469	14.5	80.6	79.2-82.0	91	2.8	104	3.2	78.4	77.0-79.8
Codes O24.4-O24.9	482	14.9	474	14.6	80.8	79.4-82.2	91	2.8	99	3.1	79.5	78.1-80.9
Hypertensive disorders												
Previous or during pregnancy	213	6.6	239	7.4	69.5	67.9-71.1	73	2.3	47	1.4	77.9	76.5-79.3
Moderate or severe pre-eclampsia	96	3.0	106	3.3	70.8	69.2-72.4	31	1.0	21	0.6	78.1	76.7-79.5
Eclampsia	2	0.1	8	0.2	12.5	11.4-13.6	7	0.2	1	0.0	50.0	48.3-51.7
Premature labor*** (O60.0)	141	4.3	87	2.7	57.5	55.8-59.2	37	1.1	91	2.8	35.5	33.9-37.1
Premature labor***§	141	4.3	326	10.0	32.5	30.9-34.1	220	6.8	35	1.1	75.2	73.7-76.7
Premature delivery	434	13.4	295	9.1	85.8	84.6-87.0	42	1.3	181	5.6	58.3	56.6-60.0
Placental abruption	41	1.3	40	1.2	75.0	73.5-76.5	10	0.3	11	0.3	73.2	71.7-74.7
Delivery												
Type of pregnancy												
Singleton	3,134	96.5	3,179	97.9	98.3	97.9-98.7	54	1.7	9	0.3	99.7	99.5-99.9
Twin	105	3.2	106	3.3	94.3	93.5-95.1	6	0.2	5	0.2	95.2	94.5-95.9
Triple	5	0.2	8	0.2	50.0	48.3-51.7	4	0.1	1	0.0	80.0	78.6-81.4
Delivery												
Vaginal	2,508	77.3	2,51	77.3	99.5	99.3-99.7	13	0.4	11	0.3	99.6	99.4-99.8
Operative	343	10.6	373	11.5	88.2	87.1-89.3	44	1.4	14	0.4	95.9	95.2-96.6
Caesarean	731	22.5	738	22.7	98.5	98.1-98.9	11	0.3	4	0.1	99.5	99.3-99.7
Emergency caesarean	450	13.9	516	15.9	83.0	81.7-84.3	88	2.7	22	0.7	95.1	94.4-95.8
Planned caesarean	238	7.3	250	7.7	80.4	79.0-81.8	49	1.5	37	1.1	84.5	83.3-85.7
Episiotomy***	350	14.0	334	13.4	90.1	88.9-91.3	33	1.3	49	2.0	86.0	84.6-87.4
Perineal tears***	1,231	49.3	1,092	43.7	86.2	84.8-87.6	151	6.0	290	11.6	76.4	74.7-78.1
Postpartum hemorrhage*** (PPH)												
Immediate (O72.0, O72.1)	286	8.8	278	8.6	77.7	76.3-79.1	62	1.9	70	2.2	75.5	74.0-77.0
Diagnosis codes + manual removal of the placenta*	191	5.9	31	1.0	80.7	79.3-82.1	6	0.2	166	5.1	13.1	11.9-14.3
Severe PPH												
Relevant advanced interventional procedures	120	3.7	143	4.4	67.8	66.2-69.4	46	1.4	23	0.7	80.8	79.4-82.2
Relevant or general advanced interventional procedures	120	3.7	158	4.9	68.4	66.8-70.0	50	1.5	12	0.4	90.0	89.0-91.0
Medical abortion	153	4.7	153	4.7	91.5	90.5-92.5	13	0.4	13	0.4	91.5	90.5-92.5
Medical abortion, fetal pathology	137	4.2	117	3.6	88.9	87.8-90.0	13	0.4	33	1.0	75.9	74.4-77.4
Stillbirth												
Relevant Z37 codes¶, or O36.4 or O31.2	239	7.4	255	7.9	89.4	88.3-90.5	27	0.8	11	0.3	95.4	94.7-96.1
Relevant Z37 codes + P95	239	7.4	391	12.0	60.6	58.9-62.3	154	4.7	2	0.1	99.2	98.9-99.5
Transfer in utero	61	1.9	30	0.9	56.7	55.0-58.4	13	0.4	44	1.4	27.9	26.4-29.4

PPV (Positive predictive value), FP (false positives), FN (false negatives), CI (confidence interval); * Missing Data (MD) = 309; ** MD=24; *** MD=30

§ codes O60.0, O60.1, O60.2, O47.0; ¶ codes Z37.1, Z37.3, Z37.4, Z37.6, Z37.7

The pregnancy-related disorders are presented in Table 1. The concordance rate of gestational age at delivery was 91.8% [90.9-92.7]. Rounding up or down to the nearest whole number of WG (< 1 week), the concordance rate increased to 98.3%.

The indicators of delivery are presented in Table 1. Regardless of the algorithm explored, the PPV for vaginal delivery was over 99%. In order to select severe postpartum hemorrhage, we explored advanced interventional procedures which indicated a second-line therapy (arterial embolization, uterine or hypogastric artery ligation, hemostasis hysterectomy).

The concordance rate between the vital status and the diagnosis codes for stillbirth from newborn discharge abstracts was 95.4% [94.7-96.1] for singleton pregnancy. For multiple pregnancies, the rate was 99.7% for the first- or the second-born child, and 100% for the third-born child.

3.2. Newborn indicators

The concordance rate of newborn weight was 91.3% [90.3-92.3] in singleton pregnancy. The rate was 79.1% [70.5-87.5] for first- and second-born in cases of multiple pregnancies. As regards the first-born child, the median gap between the newborn weight mentioned in the medical record and the weight specified in the discharge abstract was 100g.

4. Discussion

4.1. Main findings

Our study has shown that the French hospital perinatal discharge abstracts database, when compared with hospital medical records (our gold standard), regarding parturient women, pregnancy and delivery is highly accurate. Gestational diabetes had a higher metrological quality than the other types of diabetes. Several algorithms were studied using the data of hospital stay during pregnancy or during delivery. They did not improve the identification of women who had gestational diabetes. The algorithm including one of two gestational diabetes diagnosis codes (O24.4, O24.9) from the hospital stay of delivery remains the most performant. The algorithms for stillbirths and termination of pregnancy for medical reasons were found to be accurate.

4.2. Strengths and limitations

Our national study included maternity units of all types and all volumes of deliveries. Moreover, in France almost all deliveries occur in a hospital. We explored a large number of perinatal indicators: 70 items were collected for each hospital stay by a single technician specialized in clinical studies. Some limitations also have to be acknowledged. The data collection was performed by a technician, perfectly trained, but not by a health professional. The maternity units included in our study were geographically not distributed uniformly throughout the French territory. It is very important to assess the data quality in case-finding algorithms so that the design of an envisaged study can be adjusted according to the quality of data (4,8). Our results suggest that it seems possible and worthwhile to conduct studies on women with a history of gestational diabetes. The use of a complementary French database for ambulatory care (treatment, biology) has been suggested as a means to identify women with gestational diabetes or pre-existing diabetes, but it is necessary to assess the data

quality of this algorithm (9,10). The quality of the French hospital database for stillbirth and medically indicated abortion was high. Thus, it seems possible to use these two indicators not only for descriptive studies, but also for longitudinal studies.

Conclusion

This first national validation study of a large set of perinatal algorithms has shown that French hospital database is an appropriate data source for subsequent epidemiological studies. The choice of the algorithm may vary, depending on the aim of the study. Nevertheless, the data quality of the algorithms should be taken into account to define the design of the study.

References

- [1] Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* oct 2007;10(5):326-35.
- [2] Quantin C, Cottenet J, Vuagnat A, Prunet C, Mouquet M-C, Fresson J, et al. [Quality of perinatal statistics from hospital discharge data: comparison with civil registration and the 2010 National Perinatal Survey]. *J Gynecol Obstet Biol Reprod (Paris).* nov 2014;43(9):680-90.
- [3] Remontet L, Mitton N, Couris CM, Iwaz J, Gomez F, Olive F, et al. Is it possible to estimate the incidence of breast cancer from medico-administrative databases? *Eur J Epidemiol.* 2008;23(10):681-8.
- [4] Quantin C, Benzenine E, Hägi M, Auverlot B, Abrahamowicz M, Cottenet J, et al. Estimation of national colorectal-cancer incidence using claims databases. *J Cancer Epidemiol.* 2012;2012:298369.
- [5] Couris CM, Polazzi S, Olive F, Remontet L, Bossard N, Gomez F, et al. Breast cancer incidence using administrative data: correction with sensitivity and specificity. *J Clin Epidemiol.* juin 2009;62(6):660-6.
- [6] Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol.* août 2011;64(8):821-9.
- [7] De Coster C, Quan H, Finlayson A, Gao M, Halfon P, Humphries KH, et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. *BMC Health Serv Res.* 15 juin 2006;6:77.
- [8] Wentzell N, Schink T, Haug U, Ulrich S, Niemeyer M, Mikolajczyk R. Optimizing an algorithm for the identification and classification of pregnancy outcomes in German claims data. *Pharmacoepidemiol Drug Saf.* 2018;27(9):1005-10.
- [9] Fosse-Edorh S, Rigou A, Morin S, Fezeu L, Mandereau-Bruno L, Fagot-Campagna A. [Algorithms based on medico-administrative data in the field of endocrine, nutritional and metabolic diseases, especially diabetes]. *Rev Epidemiol Sante Publique.* oct 2017;65 Suppl 4:S168-73.
- [10] Billionnet C, Mitanchez D, Weill A, Nizard J, Alla F, Hartemann A, et al. Gestational diabetes and adverse perinatal outcomes from 716,152 births in France in 2012. *Diabetologia.* 2017;60(4):636-44.

Exploring Patient Path Through Sankey Diagram: A Proof of Concept

Antoine LAMER^{a,b,1}, Gery LAURENT^a, Sylvia PELAYO^a, Mehdi EL AMRANI^c,
Emmanuel CHAZARD^b and Romaric MARCILLY^a

^aUniv. Lille, INSERM, CHU Lille, CIC-IT/Evalab 1403 - Centre d'Investigation clinique,
EA 2694, F-59000 Lille, France

^bUniv. Lille, CHU Lille, Public Health Department, EA 2694, F-59000 Lille, France

^cUniv. Lille, CHU Lille, Department of Digestive Surgery and Transplantation, F- 59000
Lille, France

Abstract. Managers, physicians and researchers need to study patient's path for purposes of management, quality of care and research. We present the proof of concept of the use of a flow diagram, the Sankey diagram, to visualize the trajectory of a population that experienced an event. This representation was tested with two case studies in populations from the anesthesia data warehouse of Lille University Hospital. For the 551 patients undergoing a pancreaticoduodenectomy, Sankey diagram helped us identify atypical care paths of patient being transferred too late in an intensive care unit. For 473953 patients who have had anesthesia procedure, Sankey diagram highlighted that mortality and re-operation rates increase with the number of operations. This preliminary work has been well received by end-users and allowed managers, physicians and researchers to visualize the paths of patients and to provide visualization support for research questions. This work will be followed by generalization.

Keywords. Data Visualization, Data Reuse, Patient Path, Sankey Diagram

1. Introduction

To display complex information in a most effective way, graphical representations are more appropriate than tables and figures [1, 2]. In healthcare facilities, to have access to visual representations of patients' path would be of great value for assessing quality of care (e.g. to detect atypical care paths) and for management purpose [3]. From a research perspective, studying patient flows may generate hypotheses about patient care that can be assessed later with classical biostatistics methods.

Designing and providing clear and efficient visualization of patients' paths faces several issues:

- Hospital Information Systems (HIS) generate a huge amount of data and the patient path is documented by many variables (admission and discharges modes, type of units, succession of interventions and re-interventions, medical acts, diagnostics, etc). This makes it difficult to extract relevant information;

¹ Corresponding Author, Antoine Lamer, 2 avenue Oscar Lambret, 59037 Lille cedex, France; E-mail: antoine.lamer@chru-lille.fr.

- Patients' stays differ in complexity and may represent a large number of patterns (e.g. one patient went through two care units and underwent one operation while another went through 6 units and underwent two operations).

Several kinds of diagrams are commonly used to represent flows. From a practical point of view, a flow is a succession of steps, each one defined by an initial state, a final state, and a quantity. From a graphical point of view, it is represented by links carrying information about quantity from start nodes to end nodes. Table 1 compares 8 types of flows diagrams [4]. Most of these representations enable to visualize only one step of a process. On the contrary, the Sankey diagram enables to represent several steps along with the information about flows' volumes, and to manage the complexity of the patient paths.

This work aims to assess whether the Sankey diagram is feasible to represent patient paths, in a way that would be readable and understandable by clinicians. With this goal in mind, two cases were used: (i) paths of patients with the pancreaticoduodenectomy and (ii) association between re-operation and mortality.

Table 1. Comparison of flow diagrams

Flow diagram	Advantage	Limit
Network Diagram highlights connections between nodes at a given step.	The connections may be directed and weighted.	It addresses only one step of the whole path.
Chord Diagram is similar to network diagram, except that nodes are represented in a circle.	The same as Network diagram.	It addresses only one step of the whole path and presents reading difficulties.
Dendrogram is a tree displaying paths from the first to last step and considering the intermediate states.	It displays successive steps of the path with detail about the count.	It displays as many final nodes as paths, which overwhelms the diagram.
Sunburst Diagram is a circular representation of the dendrogram: the inner circle represents the first step of the process and each additional layer represents a new step.	It displays successive steps of the path with detail about the count, and aggregate links at every steps.	The number of steps must be limited. At each step, patients with the same criteria are displayed on different parts of the circle, which makes the reading harder.
Sankey Diagram displays flows between successive steps through arrows whose thickness is proportional to flow quantity.	It displays successive steps of the path with detail about the count.	For clarity sake, the number of steps must be limited.
In the Agenda Diagram , each row represents an individual and each cell represents a time unit.	It displays every detail for each patient.	It does not synthesize information on the whole population.
Connection Map is connecting locations placed on a map.	It highlights flows between distant places.	It represents only one step of the path and does not consider the volume of flows and the temporal sequence.
Funnel Chart represents the successive proportions of a given population that meet given criteria.	Readability.	Only one final sub-population.

2. Materiel and methods

We worked on two different cases to illustrate the design of the Sankey diagram used to represent patient paths. Each case aimed to answer a question: (i) in the digestive surgery department, what is the patient volume in each care unit after surgical operation for pancreaticoduodenectomy and what are the atypical patient paths? (ii) how mortality is related to the number of surgical operations during an hospital stay?

To design the Sankey diagrams and representing answers to those questions, we used data from the anesthesia data warehouse of Lille University Hospital. This data warehouse stores and organizes data about surgery and anesthesia procedures (drugs, vital signs, steps of surgery) with administrative data (hospital stay characteristics, medical unit, entry and discharge modes, medical and surgical acts, diagnoses) [5,6].

The Sankey diagrams were designed following three stages. The first stage is crucial and consists in determining information needed to answer the question. The information are summarized in Table 2.

Table 2. Aggregated data to upload in the graphical library

Source node	Case study 1	Case study 2
Population	Patients who undergo a duodenopancreatectomy (HNFA007 code)	Patients who undergo an operation
Event under scrutiny	Admission to unplanned critical care unit after operation	Death and re-operation rates
Steps of the path	Successive medical units after operation	Discharge status after each operation

The second stage consists in aggregating the data and compute the count of patients in each step between two successive states of the path. For each link, we had to provide a source node, a target node and the quantity between the two nodes. Finally, these aggregated data are loaded into graphical libraries (D3.js and D3sankey [7]).

The resulting Sankey diagrams were presented to clinicians to get their feedback on the readability, the understandability of this visual representation and its impact.

3. Results

3.1 Case study 1 : Patient flow for pancreaticoduodenectomy

Operations of pancreaticoduodenectomy were selected based on the code HNFA007 of the French medical act classification (CCAM, Classification Commune des Actes Médicaux). For operations between 2010 and 2018, 551 patients were included. We selected two significant states of the hospital stay for pancreaticoduodenectomy: the post-operative unit and the last unit before discharge from the hospital. These two states were characterized with the type of unit (conventional, continuous care or intensive care unit) and final discharge status (mortality) was reported. The Figure 2 is the Sankey representation for this population.

Two managers and two physicians examined the Sankey diagram. They were able to describe patient flows in each unit after operation. Moreover, physicians highlighted atypical paths: some patients were transferred from conventional care unit to critical care after operation, and some patient died even though they did not go through critical care beforehand. They reported that this Sankey diagram was intuitive to read, more understandable than a table but that they needed contextual information when the graph is accessed for the first time. Physicians will now investigate why these patients were not treated earlier in an appropriate care unit.

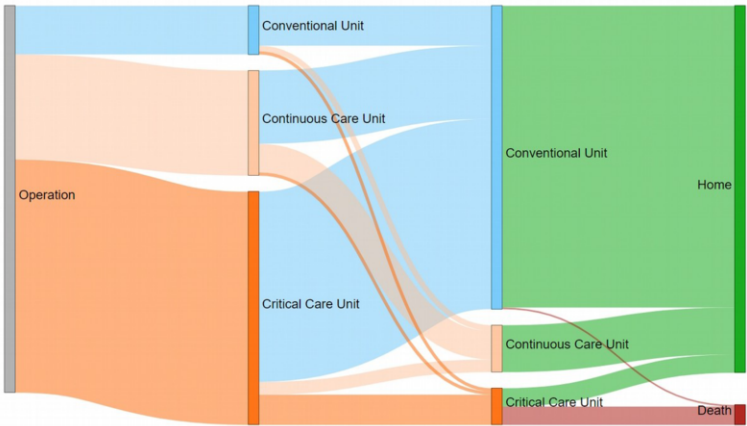


Figure 1. Patient flow for pancreaticoduodenectomy

3.2 Case study 2: Mortality rates after re-operation

Between 2010 and 2018, 473953 patients underwent operation. For each patient, we computed the number of operations, and final discharge status (death). In order to get the most out of the diagram, two key points were relevant. First, we had to define a maximum number of operations to be covered. Secondly, we had to display results in percentages to compare modalities with small numbers (3rd and 4th surgical procedures) with larger modalities (1st and 2nd surgical procedures). With the Sankey diagram in Figure 2, we can see that mortality and re-operation rates increase with the number of operations performed during the same hospital stay. This diagram has been shown to 2 managers and 2 physicians, and they all agreed on the intuitiveness of the information displayed. They only asked for contextual information to know the source of the data. According to them, this result will be the starting point of research program about how going through multiple operations could impact mortality rates.

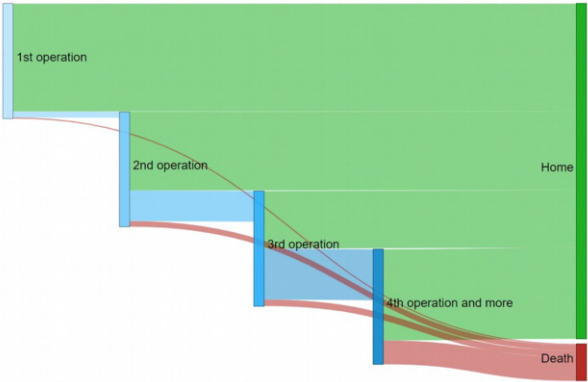


Figure 2. Sankey diagram representing the mortality and re-operation rates according to the number of operations during a hospital stay

4. Discussion

In this paper, we offer a way of visualizing patients' path with the Sankey diagram. This representation allowed managers, physicians and researchers to visualize the paths of patients, to detect atypical patient flows between units and to provide visualization support for research questions. We will now investigate which signals in operating room and post-anesthesia care unit could be used to detect patients prone to complications and refer them to an intensive care unit as soon as possible.

Usually, end-users are provided with data in a table format, on a case-by-case basis, or for a single step of the patient path. With the Sankey diagram, the benefits reported by users are that this representation is easily understandable, gives a useful overview of patient path and offers the opportunity to identify atypical paths that they may not have seen otherwise. It could also be used as a way to validate their hypotheses and further investigate.

However, the implementation of such diagrams might be challenging as patient paths have more than two steps and are complex. Thus, the data aggregation must be adapted to the specific issue to deal with. Moreover, the steps displayed must be selected to highlight important facts and avoid overload of information.

A limitation of the Sankey diagram, with its current implementation, is that it is not possible to follow an individual path from the first to the last node. The next step is to improve the current Sankey representation by adding a traceable multi-level feature. This would represent a more precise way to highlight a specific path.

Another perspective would be to offer an online Sankey visualisation tool, in which the users would select the targeted population (e.g. a surgical act) and predefined steps (e.g. surgical complication) to have a display of the corresponding patients' path.

References

- [1] West VL, Borland D, Hammond WE. Innovative information visualization of electronic health record data: a systematic review. *J Am Med Inform Assoc JAMIA*. mars 2015;22(2):330-9.
- [2] Few S. *Information Dashboard Design : The Effective Visual Communication of Data* / S. Few. 1 janv 2006;
- [3] Chittaro L. Information visualization and its application to medicine. *Artif Intell Med*. mai 2001;22(2):81-8.
- [4] Holtz Y, Healy C. From data to Viz | Find the graphic you need [Internet]. [cited on 24th sept 2019].
- [5] Available on : data-to-viz.com
- [6] Lamer A, Jeanne M, Vallet B, Ditleyeu G, Delaby F, Tavernier B, et al. Development of an anesthesia data warehouse: Preliminary results. *IRBM*. 2013 Dec;34(6):376-8.
- [7] A. Lamer, G. Ficheur, L. Rousselet, M. van Berleere, E. Chazard, A. Caron, From Data Extraction to Analysis: Proposal of a Methodology to Optimize Hospital Data Reuse Process, *Stud. Health Technol. Inform* 247 (2018), 41-45.
- [8] Bostock M. D3.js - Data-Driven Documents [Internet]. [cited on 9th July 2019]. Available on: <https://d3js.org/>
- [9] CCAM en ligne - CCAM [Internet]. [cited on 8th January 2019]. Available on: <http://www.ameli.fr/accueil-de-la-ccam/index.php>

Extending Contsys Standard with Social Care Concepts: A Methodology Proposed by the UNINFO Working Group in Italy

Fabrizio PECORARO^{a,1}, Daniela LUZIA, Elaheh POURABBAS^b, Fabrizio L. RICCI^{a,b}
and Angelo ROSSI MORI^a

^a*Institute for Research on Population and Social Policies, National Research Council, Italy*

^b*Institute for System Analysis and Computer Science “Antonio Ruberti”, National Research Council, Italy*

Abstract. The increasing demand for territorial services requires the improvement of the coordination and cooperation among stakeholders in planning and delivery of integrated health and social services. In this scenario, to improve the communication among stakeholders there is a need of a formal conceptual model that facilitates the interoperability between organizations and professionals. This paper presents the methodology adopted by a UNINFO working group established in Italy to extend the ContSys standard with social care concepts to integrate health and social care contexts in a continuity of care perspective. An example of this extension is also provided considering the definition of patient's care plans.

Keywords. Continuity of Care, Integrated Care, ContSys standard, Semantic interoperability, Health and social care services

1. Introduction

In recent decades there has been a change in the provision of health services privileging home care over hospitalization [1]. This is considered a sustainable approach that not only leads to significant cost savings, but also allows optimal results to be achieved in providing high-quality health services as well as offering an improvement in the individual's quality of life [2]. Furthermore, the rapid development of medical and non-medical technologies facilitates the provision of home care services, supports care coordination among professionals as well as allows remote patient monitoring [3]. However, a critical aspect considering the provision of services both in the home environment and in the community is their fragmentation that leads to scarce outcomes and wasted resources [4]. Thus, in this perspective, an important target to be accomplished is the improvement of the coordination and cooperation among stakeholders considering the different needs of the target population both between primary and secondary care (vertical perspective) and between health and social care settings (horizontal perspective). In this scenario, one of the first issues to be tackled to

¹ Fabrizio Pecoraro, Institute for Research on Population and Social Policies, Via Palestro, 32 – 00185 – Rome, Italy; E-mail: f.pecoraro@irpps.cnr.it.

improve the communication among stakeholders is the definition of a formal conceptual model that defines the concepts needed to facilitate the interoperability between organizations and professionals. Among the standards developed in healthcare ContSys (A system of concepts for the continuity of care, formally ISO 13940 [5]) represents a suitable starting point to accomplish our task. It is focused on semantic interoperability that is one of the major requirements to achieve the continuity of care. ContSys has the advantage of considering continuity as the integration of different health services provided by formal and informal carers belonging to different types of organizations such as primary care, hospital, territorial, community. Moreover, this standard provides a comprehensive conceptual framework to describe generic concepts that represent both the content and context of the health services with well-defined relationships among a large set of classes that describe in details the healthcare process [6]. However, even if ContSys states that also social care has the objective to influence, restore and maintain health, this standard only indirectly include social care within the wider spectrum of health concepts. This paper starts from a previous work [7,8] where we introduced a preliminary idea of an integrated model that collates healthcare, social care and assistive domotics elements to define the conceptual model of an interoperable open platform, called H@H, that organizes the provision of a set of home care services. The resulting conceptual model was confined on the portion of the ContSys standard fitting with the purposes of the H@H project. Starting from the results achieved in these papers a working group was launched at the UNINFO, the branch of the Italian National Unification (UNI) Standardization body that tackles the Information and Communications Technology (ICT) activities, to put in practice the idea of extending the ContSys standard as a whole.

The aim of this paper is to describe the methodology adopted by the working group to extend the ContSys standard with social care concepts to establish an overall schema that models the integration between care professionals in a continuity of care perspective. The steps of the methodology are described in Section 2, while Section 3 reports the application of these steps to define patient's care plans.

2. Methodology

The extension of the ContSys model was performed on the basis of the following steps.

Step 1. Analysis of the ContSys concepts in the light of the social context. Given the structure of the ContSys standard and the relevant technical document, each clause of the conceptual schema was analysed taking into account its concepts as well as the related classes. Each concept was analysed considering its definition, the relationships with the relevant classes (type, name and cardinality) as well as notes and examples. On the basis of this analysis the working group determined whether the concept can be:

- a. *Included as it is.* This category refers to universal concepts that already include in their definition, notes and examples references to social care. Examples are: subject of care, health resources such as funds;
- b. *Included but extending the vocabulary with social references.* These concepts need to be revised in terms of notes and/or examples or slightly modified in their definition. Examples are: health matter, contact.
- c. *Included identifying a new specific concept.* These concepts explicitly describe the clinical process but have a counterpart within the social context and can be integrated within the model. This required the creation of a new concept

targeted to the social model. This can imply the redefinition or the introduction of new relationships. Examples are: clinical process, health issue.

- d. *Excluded* given that these concepts describe the clinical side of care and are not needed for the social context, such as adverse event.

Step 2. Definition of the combined model. Once all concepts were analysed a conceptual schema describing the social side of care and its relationships with the clinical side of care was defined. The resulting model was composed of three main sections. The first two sections describe separately the healthcare and the social care contexts, while the third one concerns concepts that are shared between the two above mentioned sections. Moreover, additional associations are determined to capture the relationships between the health and the social parts of the schema.

Step 3. Definition of the integrated model. Once the model that combines both the healthcare and the social care perspectives is defined, the final step of the methodology concerned the identification and modelling of the part of the model where services are delivered in a professional collaboration and integration perspective. This is mainly applied considering the assessment and planning activities that can be carried out by a multidisciplinary team of specialists belonging to health and social care organizations.

The analysis carried out during each step of the methodology can facilitate the refinement of the models defined in the previous steps in an iterative approach. In the following paragraph an example of the application of the methodology is described analysing the healthcare planning clause of the ContSys standard [5].

3. Results

Figure 1 shows the combined model (see step 2 of the methodology described in the previous paragraph).

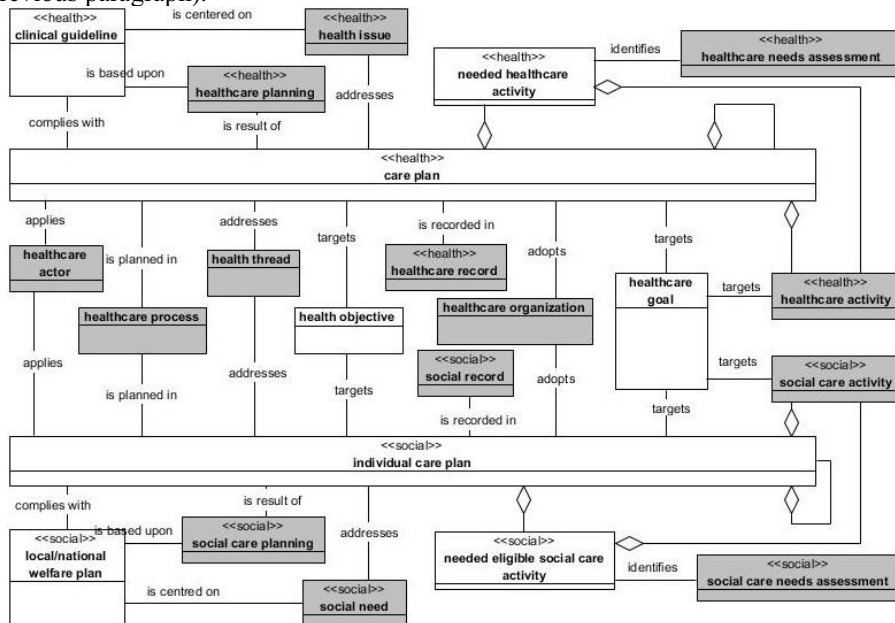


Figure 1. Combined model highlighting classes defined in the ContSys Healthcare Planning clause (white) and those defined in other clauses (grey). <<Stereotype>> highlights health and social care classes.

The core class of the social section is the Individual Care Plan (ICP) (also known as Piano Assistenziale Individuale (PAI) in Italy) that is focused on one or more social needs. It is composed of a set of needed social care activities and a set of planned or delivered social care activities. The plan is stored in a social record that concerns a specific subject of care. A similar description can be seen considering the healthcare section of the schema, while in the common section a set of goals and objectives that are in common in both contexts are associated with health and social care plans. Similarly, the health thread, that is in common with the two sections of the schema, is adopted to associate both health issues and social needs of the relevant subject of care. As introduced in the methodology, the integration of health and social care contexts does not only imply that a subject of care accesses to these services separately, but also to a set of activities that are carried out by a multidisciplinary team in an integrated approach in a continuity of care perspective. This may occur especially during the planning activities. In case of an integrated context, a set of activities may be provided by a team on the basis of a collaborative and continuity of care approach. For this reason, the planning clause of the model was additionally extended by including the integrated plan of care concept. It is important to stress that an integrated plan cannot be considered only as a composition of health and social care plans because it has to:

- Incorporate a set of needed and provided *health, social care and integrated care activities*;
- Target a set of *objectives and goals*, that are shared between contexts;
- Be stored in an *integrated care record*;
- Be applied by different *actors* involved in the health and social care sectors;
- Plan a *healthcare process* that describes both health and social care contexts;
- Address a *health thread* that links both *health issues* and *social needs*;
- Be planned during a set of integrated activities involving both health and social care professionals (*integrated planning*);

On the basis of these features an additional model that describes the healthcare integrated planning clause is defined and shown in Figure 2.

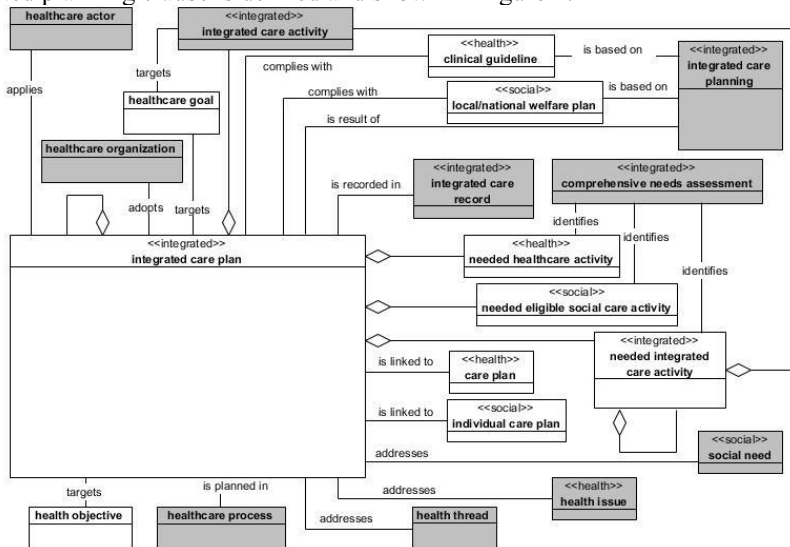


Figure 2. Integrated model highlighting classes defined in the ContSys Planning clause (white) and those defined in other clauses (grey). <<Stereotype>> highlights health, social and integrated care classes.

4. Discussion

This paper describes the methodology proposed by the UNINFO Working Group in Italy to extend the ContSys standard to social concepts. It provides an example of the definition of the individual care plan established in the social context to track the activities (both needed and provided) related to one or more social needs of a specific subject of care. Moreover, as different services can be planned and provided by a multidisciplinary team composed of both health and social care professionals, an integrated plan is envisioned. The adoption of the integrated model could facilitate the interoperability between systems as well as the coordination and cooperation activities of the different actors involved in the provision of services in a comprehensive continuity of care process.

At the moment the UNINFO team has analyzed the different concepts of the standard identifying an extension of the model as well as capturing the part of the model where an integrated view of the care can be modelled. Moreover, thanks to the interdisciplinary composition of the team, that also comprises researchers and professionals belonging to social care, a set of scenarios have been collected to capture and assess the completeness of the extended model. The debate set up within the UNINFO group can be a starting point to provide suggestions to the established ISO working group for the revision of the ContSys standard.

Acknowledgment

The authors would like to thank all the other participants at the UNINFO working group: Pierangelo Sottile (UNINFO), Daniele Babusci, Antonella Ciocia (CNR-IRPPS), Giorgio Cangioli (HL7 Europe), Mario Ciampi (CNR-ICAR), Fabrizio Clemente (CNR-IC), Doriana De Marco (Italian Data Protection Authority), Davide Guerri (Dedalus), Danilo Pani (University of Cagliari) and Gregorio Mercurio.

References

- [1] H. Vaartio-Rajalin, L. Fagerström, Professional care at home: Patient-centeredness, interprofessionalism and effectiveness? A scoping review, *Health and Social Care in the Community* 27 (2019), 1–9.
- [2] L. Di Pollina, I. Guessous, V. Petoud, C. Combescure, B. Buchs, P. Schaller, M. Kossovsky, J.M. Gaspoz, Integrated care at home reduces unnecessary hospitalizations of community-dwelling frail older adults: a prospective controlled trial. *BMC Geriatrics* 17 (2017), 53.
- [3] R. Tarricone, A.D. Tsouros. Home care in Europe: the solid facts. WHO Regional Office Europe, 2008.
- [4] World Health Organization. Integrated care for older people: realigning primary health care to respond to population ageing. No. WHO/HIS/SDS/2018.44. World Health Organization, 2018.
- [5] ISO 13940, Health Informatics – System of Concepts to Support Continuity of Care, 2015.
- [6] CEN TC251, Towards concurrent use of ContSys, 13606, and HISA, Report from 2nd Workshop, Madrid, 2013.
- [7] F. Pecoraro, D. Luzi, E. Pourabbas, F.L. Ricci, A conceptual model for integrating social and health care services at home: the H@H project. Proceedings of the 18th International conference on E-Health, Networking, Application and Services, 2016. DOI: 10.1109/HealthCom.2016.77494.
- [8] F. Pecoraro, D. Luzi, F.L. Ricci, An integrated model to capture the provision of health and social care services based on the ContSys and FHIR standards, *European Journal of Biomedical Informatics* 13 (2017), 17–26.

Feasibility of Using EN 13606 Clinical Archetypes for Defining Computable Phenotypes

Archana TAPURIA,^a Dipak KALRA,^b and Vasa CURCIN^c

^a*Research Fellow, King's College London (KCL)*

^b*Professor in Health Informatics, University College London*

^c*Lecturer in Health Informatics, KCL*

Abstract. Introduction: Computable phenotypes are gaining importance as structured and reproducible method of using electronic health data to identify people with certain clinical conditions. A formal standard is not available for defining and formally representing phenotyping algorithms. In this paper, we have tried to build a formal representation of such phenotyping algorithm. **Methods:** We built EN 13606 EHR standard for building clinical archetypes to represent the computable phenotyping algorithm for 'diagnosis of cardiac failure'. As part of this work, we created a set of new clinical archetypes for defining 'cardiac failure diagnosis'. The EN13606 editor called Object Dictionary Client was used which was in-house developed by University College London. We evaluated the ability of EN 13606 to provide clinical archetypes to define EHR phenotyping algorithms using the predefined desiderata for the purpose [Mo et al]. **Results:** EN 13606 archetypes could represent phenotype components grouped and nested based on their logical meaning. It was possible to build the EHR phenotyping algorithm with the clinical elements and their interrelationships along with hierarchical structure and temporal criteria. But the specific mathematical calculation and temporal relations involved in the algorithm was difficult to incorporate. These will need to be coded and integrated within the clinical information system. These archetypes can be mapped for comparison with the openEHR models. Binding to external clinical terminology is fully supported. However, it does not satisfy all the desiderata defined by Mo et al. A possible way could be an approach using phenotype ontologies and its architectural representation integrated with ISO interoperability. **Conclusion:** The EN13606 archetypes can be used to define the phenotype algorithm that basically identifies patients by a set of clinical characteristics in their records. Phenotype representations defined in EN 13606 do not satisfy all the desiderata proposed by Mo et al. and thus currently has a limited ability to define the computable phenotyping algorithms. Further work is required to make the EN13606 standard to fully support the objective.

Keywords. computable phenotypes, EN13606, archetypes, EHRs, algorithms.

1. Introduction

A computable phenotype refers to a clinical condition or characteristic that can be determined with a computerised query to an Electronic Health Record (EHR) system or clinical data repository using a defined set of data elements and logical expressions (<https://rethinkingclinicaltrials.org/resources/ehr-phenotyping/>). Standardised

computable phenotypes can promote efficient and reliable recruitment in clinical trials across various clinical faculties and health systems.

A formal standard is not available for defining and formally representing phenotyping algorithms in a standardized way. Hence the automatic or computerized translation of phenotyping algorithms to computational implementation code is not possible. In this paper, we have tried to build a formal representation of such phenotyping algorithm using EN 13606 EHR standard. Similar work has been done to evaluate openEHR standard for building computable phenotyping algorithms [1]. This attempt shall add to the learning process.

1.1 CEN/ISO EN13606 and Clinical archetypes:

The CEN/ISO EN13606 is a European norm from the European Committee for Standardization also approved as an international ISO standard [2] intended to support the interoperability of systems and components that need to communicate EHR data: EN13606 follows an innovative Dual Model architecture. The former is structured through a reference model (RM) [3] that is an object-oriented model used to represent the generic and stable properties of health record information and the later is based on clinical archetypes [4,5]. Figure 1 shows the building blocks or classes of EN13606 RM, such as Folder, Composition, Section, Entry, Cluster and Element.

An archetype is a formal and standard way of representing the clinical data structures and their inter-relationships for a given clinical domain. They are content specifications expressed in terms of constraints on a specific reference model [6]. Archetypes specify EHR clinical data hierarchies and the kinds of data values used within each entry [7].

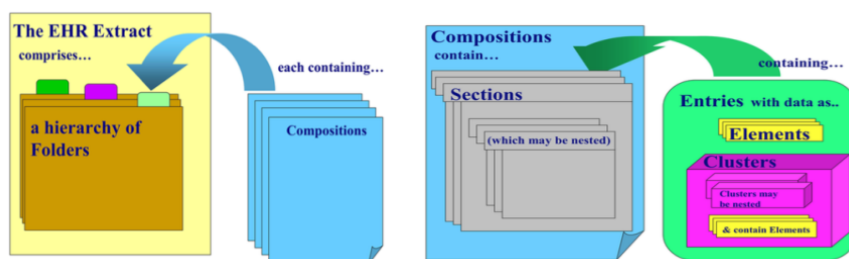


Figure 1: Structure of the reference model of EN13606 and EHR extract hierarchy [8]

2. Methods

2.1. Clinical condition/characteristic and its algorithm chosen for phenotyping:

We chose ‘cardiac failure’ as the clinical condition, diagnosis of which would need a computable phenotyping algorithm. We used a previously validated EHR phenotyping algorithm for identifying and diagnosing patients with cardiac failure. (<https://www.mdcalc.com/framingham-heart-failure-diagnostic-criteria>) The algorithm as shown in Fig. 2, represents clinical information that contributes towards the diagnosis

of cardiac failure. The algorithm was designed to identify patients with cardiac failure. Identification of various types of cardiac failure is not taken into consideration here.

2.2. Building Clinical Archetypes:

Archetypes relevant to ‘cardiac failure diagnosis’ were not found in the publicly available archetype repositories or libraries like the ‘Clinical Knowledge Manager’ by the openEHR foundation (<https://www.openehr.org/ckm/>) or ‘Ocean Informatics’ (https://www.openehr.org/industry_partners/ocean_informatics). As part of this work, we created a set of new archetypes for defining ‘cardiac failure diagnosis’ underpinned by the EHR standard EN 13606. The EN13606 editor called Object Dictionary Client (ODC) was used which was in-house developed by University College London, which is recently published, and the information models built with it are open source for use [9].

The EN 13606 Archetype describes the definitive semantic model of archetypes. It defines data types, constraints, and a reference mechanism allowing one archetype to reference another (slot). Archetypes are usually built and implemented with the help of templates, that help to combine the various archetypes together in a logical manner. EN 13606 uses the Archetype Definition Language (ADL) for building the archetypes and Archetype Query Language (AQL) for querying the data.

2.3. Desiderata for assessing clinical archetypes:

We used the desiderata published by Mo et al [10] to evaluate the ability of EN 13606 to build clinical archetypes to define EHR phenotyping algorithms. Mo et al. reviewed common features for multisite phenotype algorithms and proposed 10 desired characteristics for a flexible, computable Phenotype Record Model.

3. RESULTS

a) Archetypes created for computable phenotyping algorithm: Archetypes created were organized into a single composition with subgroups composed of number of elements (Fig. 3). Each node represents a data element, which is bound to external clinical terminologies (Read Codes, ICD-10 and SNOMED CT). Additional constraints are added to each data element including, the cardinality, optionality, datatype, value range and units for a physiological quantity, internal



Figure 2. Cardiac failure EHR phenotyping algorithm for the diagnosis of Heart failure

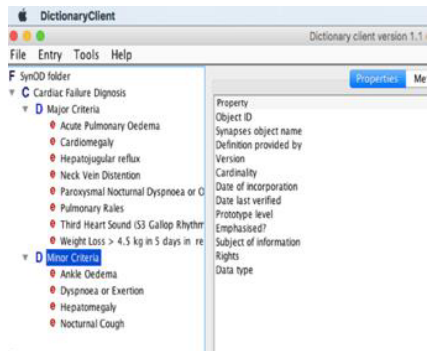


Figure 3 Clinical archetype built in ODC, an EN 13606 archetype editor

codes, etc. With EN 13606 archetypes, the components of a phenotype algorithm could be grouped or nested based on their clinical semantics into Symptoms, Signs, Investigation Results, Diagnoses, Treatment or Procedures. It can very well define the hierarchy of the data elements and the interrelationships between the data elements with the help of its Record Model (RM) properties along with the archetype slot mechanism. Still, it provides a very limited way to express the actual logic of the phenotype algorithm (especially the relational algebraic operation).

- b) Desiderata for assessing clinical archetypes: (here we have partly followed the pattern used by Papez et. Al, to represent the desiderata for assessing clinical archetypes for phenotyping algorithms) [1] Here we represent the desired desiderata [8] and where EN 13606 standard stands to satisfy it, to define phenotyping algorithms.

Table1: Summary of the feasibility of EN13606 archetypes in in defining EHR phenotyping

Desiderata by Mo. et al	Evaluation of EN 13606 EHR standard in defining computable phenotype algorithms
1. Human-readable and computable representation	Fully supported: Clinical Archetypes are defined in ADL and templates are stored as XML. EN 13606 editors can transform the archetypes (in ADL) into human-readable format such as Hypertext Markup Language (HTML) and vice versa.
2. Set operations and relational algebra	Not supported: Need to be coded and integrated with the EN 13606 clinical information model/system
3. Structured rules	Not supported: These need to be coded and integrated into the clinical information system
4. Temporal phenotype criteria	Partially supported: Archetypes support temporal criteria using points in time or time intervals. Temporal relations need to be coded within the clinical information system
5. Standardized nomenclature	Fully supported: Supports binding of the names/values of the data elements with external standardized controlled clinical terminology or classification systems (e.g. ICD-9, ICD-10, Read Codes v2/v3, SNOMED CT)
6. External interfacing	Partially supported: The entire EN13606 model can be viewed in the AWB RM schema and Class tools, and comparisons can be made with classes in the openEHR reference model.
7. Backward compatibility	Fully supported: The tracking of changes during their development and evaluation is feasible and older versions can be referred to.

4. Discussion

Since EN 13606 archetypes define the characteristics/attributes of a set of data elements and phenotyping algorithms are primarily based on identifying patients satisfying a list of criteria (e.g. signs, symptoms, investigation results, diagnosis, etc.), archetypes could be used for defining the computable phenotyping algorithms especially for their content, hierarchical structure, interrelationships of data, and temporal criteria. EN 13606 however does not satisfy all the desiderata defined by Mo et al in defining the EHR phenotyping algorithms, especially the algebraic operations and temporal relations. Natural language processing is not provided by EN 13606 and external interfacing is limited. The example of ‘cardiac failure diagnosis’ is fairly simple but more complex diagnostic criteria for various diseases would require algebraic and temporal relations to be operated and some would demand external interfacing as well.

A possible way could be an approach using phenotype ontologies and its architectural representation. The most recent published version of ISO 13606 (late 2019) refers to this approach by integrating the ISO Interoperability and Integration Reference Architecture [2] that formalizes the relationship between clinical domains and their information representation. Without that formalization, there is a risk that archetypes will proliferate to formalize pieces of a domain in disconnected and *ad hoc* ways. Ontologies can successfully formalize the representation of a domain, and therefore be an appropriate semantic anchor for the development of archetypes. With that starting point, an implementation of computable and easily searchable phenotype definitions implemented as FHIR resources is a feasible way. Furthermore, a phenotype ontology also offers rich logical relationships including the consideration of relevant context., and the desiderata defined by Mo et al. such as the delivery of structured rules, temporal criteria, etc. could be better met by the described alternative approach. By deploying the Interoperability Reference Architecture part of ISO/EN 13606, transformation between the different representation styles including archetypes can be easily performed.

5. Conclusions

EN 13606 archetypes could be used for defining the computable phenotyping algorithms especially for their contents, hierarchical structure, interrelationships of data, and temporal criteria to identify patients with a set of clinical criteria. However, it does not satisfy all the desiderata defined by Mo et al. and hence further attempts are required to achieve the objective, especially for more complex phenotyping algorithms. Clinical archetype editors underpinned by EN13606 EHR standard might need to be used along with some other programmes to make it completely feasible to represent the phenotyping algorithms. Thus, EN 13606 currently has an ability to define the computable phenotyping algorithms but has its limitations.

References

- [1] Papež V, Denaxas S, Hemingway H. Evaluating openEHR for storing computable representations of EHR phenotyping algorithms. In: 2017 IEEE 30th Int Symp Comput Med Syst. 2017. p509–514.
- [2] ISO EN 13606 Standard. <http://www.en13606.org/the-ceniso-en13606-standard> (accessed 1 Dec 2019).
- [3] ISO. BS ISO EN 13606 part 1 (No. EN 13606) (eds D Kalra & D Lloyd) 2007; 1: 1–103. [Google Scholar](#)
- [4] Austin, T, Lim, YS, Nguyen, D Design of an electronic healthcare record server based on Part 1 of ISO EN13606. J Healthc Eng 2011; 2: 143–160. <http://multiscience.metapress.com/content/121507>.
- [5] ISO. BS ISO EN 13606 part 3 (No. EN 13606) (eds D Kalra & D Lloyd) 2008; 3: 1–47. [Google Scholar](#)
- [6] Health informatics – Electronic health record communication – Part 1: Reference model (ISO 13606-1:2008). <https://www.iso.org/standard/40784.html> (accessed 1 April 2018). [Google Scholar](#)
- [7] Tapuria, A., Kalra, D., & Kobayashi, S. (2013). Contribution of Clinical Archetypes, and the Challenges, towards Achieving Semantic Interoperability for EHRs. Healthcare informatics research, 19(4), 286-92.
- [8] Tapuria, A., Bruland, P., Delaney, B., Kalra, D., & Curcin, V. (2018). Comparison and transformation between CDISC ODM and EN13606 EHR standards in connecting EHR data with clinical trial research data. DIGITAL HEALTH. <https://doi.org/10.1177/2055207618777676>
- [9] Austin, T, Sun, S, Lea, N Patterns: a simple but expressive data modelling formalism. Int J Knowledge Eng Data Mining 2016; 4(1): 74–92. [Google Scholar](#), [Crossref](#)
- [10] H. Mo, W. Thompson, L. Rasmussen, et. al, “Desiderata for clinical archetypes of EHR-driven phenotype algorithms”. JAMIA, vol. 22, 2015, pp. 1220-30, doi: 10.1093/jamia/ocv112

Acknowledgement: This research is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London.

Generation of Fine Grained Demographic Information for Epidemiological Analysis

Timo WOLTERS ^{a,1}, Oke WÜBBENHORST ^a, Christian LÜPKES ^a and
Andreas HEIN ^b

^a *Division Health, OFFIS - Institute for Information Technology, Escherweg 2,
Oldenburg, Germany*

^b *Department of Health Services Research, Carl von Ossietzky University Oldenburg,
Oldenburg, Germany*

Abstract. Cancer risks may be influenced by local exposures such as working conditions or nuclear waste repositories. To find influences, local accumulations of cancer rates are used, for which finely granulated data should be utilized. In particular, high-resolution demographic data for a reference population are important, but often not available for data protection reasons. Therefore, estimation methods are necessary to approximate small-scale demographic data as accurately as possible. This paper presents an approach to project existing epidemiological and public data to a common granularity with respect to attribute characteristics such as place of residence, age or smoking status to allow for analyses such as local accumulations and consistently falls below an average relative error of 5%.

Keywords. Health Data Science, Data Analysis, Population Synthesis, Microsimulation

1. Introduction

The risk of cancer can be increased by the degradation of nuclear waste repositories and the associated radiation exposure of the immediate environment. Local accumulations are a type of analysis in cancer epidemiology to correlate geospatial impact with cancer risk and requires finely granular data from cancer incidences and a reference population. In epidemiological cancer registries, the reference population is often more coarsely granular than the incidences due to data protection, making an approximation of the population at the incidence level desirable. This paper presents an approach for the generation of fine-grained data using iterative proportional fitting (IPF) combined with cellular automata (CA) for exemplary application in analyses of epidemiological cancer registration (ECR) in Lower Saxony and evaluates the accuracy and reliability of the generation.

In the ECR, data is evaluated pseudonymised by the registry unit, which as a result of the analysis for the verification of a suspicion can request the underlying not pseudonymised data from the confidentiality unit [1]. In order to form a suspicion, exact results are therefore not necessary at first, since the analysis is finally carried out on the

¹Corresponding Author: Timo Wolters, OFFIS - Institute for Information Technology, Escherweg 2, Oldenburg, Germany; E-mail: timo.wolters@offis.de.

data of the confidentiality unit. However, too large an approximation error leads to false suspicions and should therefore be minimized.

Incidences are present in the ECR for Lower Saxony in a geometric grid with an edge length of 1km (1km grid) while the reference population is present at municipality level. Figure 1 shows the difference between the municipality level and a 1km grid, while the actual distribution of the population is shown in a 100m grid using the municipality of Oldenburg. The 1km grid has a much finer resolution than the municipality and better reflects the spatial distribution of the population, especially with respect to the temporal stability of a grid in contrast to frequently changing administrative boundaries of municipalities [1].

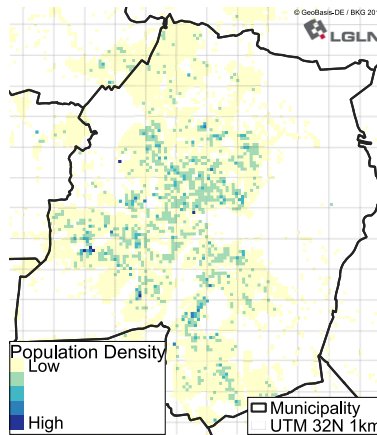


Figure 1. Granularity comparisons between a municipality, an 1km grid and an 100m grid for Oldenburg in Lower Saxony.

After Section 2 has outlined related work on population synthesis and microsimulations, Section 3 describes our approach to generate fine-granular data. Section 4 explains the implementation details and Section 5 evaluates the generation of the population, while Section 6 describes future work and summarises this contribution.

2. Related Work

Ballas et.al. investigate in [2] location-based microsimulation systems for population synthesis. Population synthesis usually combines fine granular census data with coarse granular data to approximate a synthetic population between two censuses or to add missing attributes. Population synthesis can be divided into population reconstruction, probabilistic reweighting and deterministic reweighting.

Iterative Proportional Fitting (IPF) is according to Moretti et.al. [3] a method for deterministic reweighting to map the distribution of coarse granular data (deaggregated data) to fine granular data (aggregated data). For this an unambiguous mapping between deaggregated and aggregated data is necessary. This mapping does not exist between a 1km grid and administrative boundaries of municipalities, so that IPF alone would not be sufficient to synthesize a reference population.

Basse et.al. [4] deal with exactly this problem using cellular automata (CAs) and neural networks. Cellular automata are used to simulate land use for grid cells in border regions, while neural networks are used as a probabilistic model. However, the application of neural networks requires a large demographic database, which is not available in ECR. Therefore, the combination of IPF and CA in this paper is necessary for population synthesis in order to counteract problems in border regions and insufficient data volumes.

3. Concept

For the concept, the data basis should first be considered. The aggregated data for IPF contain the demographic information with 10-year age classes and without reference to the two gender classes in a 100m grid of the 2011 census. The deaggregated data contain demographic data at municipality level with 1-year age classes with reference to the gender classes and sparsely smoker status or occupation as cancer specific attributes. Additional data are information from the census on households and families and are used in the integration of CA with IPF. From the deaggregated data, statistical ratios such as the proportion of a 1-year class in a 10-year class are extracted as additional information.

This results in a vector $U = \{U_1, U_2, \dots\}$ of attributes that are to be generated with a function $P_{U,C}$ by assigning a probability $p_{i,j}$ to each attribute characteristic $u_{i,j} \in U_i$, taking into account the deaggregated data as context C . $P_{U,C}$ gets as argument a vector $A = \{a_1, a_2, \dots\}$ of the aggregated data so that a random number $\pi \in [0; 1]$ together with A calculates a vector $u = \{u_1, u_2, \dots | u_i \in U_i\}$ representing a single person with all required attributes. A CA then checks the conformity of the generated data with the family, household and demographic data and gradually adapts the demography so that the information on families and households is reflected in the demographic data, but meets the constraints of the municipality demography.

In the application, $U = \{age, gender, residence\}$ is used with 1-year age groups up to 80 and the 100m grid of the census as a place of residence. The vector $A = \{age_{10}, residence, gender_n, households, families\}$ contains the 10-year age groups up to 80 of the census, the same place of residence and the number of men and women living in the grid cell together with the information on families and households for the CA. Other attributes such as smoker status or hormone receptor information for lung cancer or breast cancer, for example, are also possible for U and A if the necessary data are available. Based on epidemiological cancer registration, smoking status information is rare and hormone receptor information is not available, but hormone receptor information should be reported in the clinical cancer registration for estrogen in breast cancer.

After generation, an additional function \bar{p} is applied, which can serve as a bias. \bar{p} is particularly useful if generation is to be carried out for years other than the census year, since population changes and changes in administrative boundaries can be modelled here. On the other hand, this function can be used to correct inconsistencies between different data sources.

4. Implementation

The data of the census are available in tables with about 6 columns, but 61 million entries. Partitioning based on characteristics (age, gender) is essential to shorten access times

to the data. The geo references of the census are used to link the sub-data sets of the aggregated data and a mapping to municipalities is used to link the deaggregated data.

Population synthesis uses IPF for an initial statistical distribution of personal characteristics in the grid cells of the municipality. On the basis of this distribution, CA is used to generate people probabilistically until the statistical distribution is approximated, taking into account household and family data, or a certain number of iterations has been reached. However, the average number of persons per grid cell is 6, so that the empirical law of large numbers cannot be applied to the probabilistic method to approximate the statistical distribution. An adaptation of probability from [5] helps to achieve faster convergence despite small numbers of people. For this purpose, logistic growth is used as an adaptation function with the turning point $(1, p_i)$ for the basic probability of a feature, and the limits are defined by $p_u = 1.5 \cdot p_i$ and $p_l = 0.5 \cdot p_i$. The abscissa x counts the occurrence of an event and the higher x is, the greater the probability of the counter-event, whereby the occurrence of the counter-event x can either be set to 1 (hard reset) or reduced by 1 (soft reset). \bar{p} is trained with municipality data from the census year 2011 and a target year, e.g. 2017, so that the generation can also be used for other years.

5. Evaluation

The evaluation of the data quality on the 1km grid is not possible, because data with comparable accuracy are missing. Instead, the generated 100m grid cells of the map step are aggregated at the municipality level and compared with the previous municipality data for the year 2011. \bar{p} thus has no influence on the data quality, since the identity function is used for 2011. Since the procedure is probabilistic, several iterations should be considered for evaluation. The influence of probability adaptation (PA) should also be measured to determine the necessity and contribution to stability of the adaptation. A total of 12 iterations with and without probability adaptation were performed, with Figure 2 showing the best results of the iterations based on the relative error for the generated gender distribution in the municipality average.

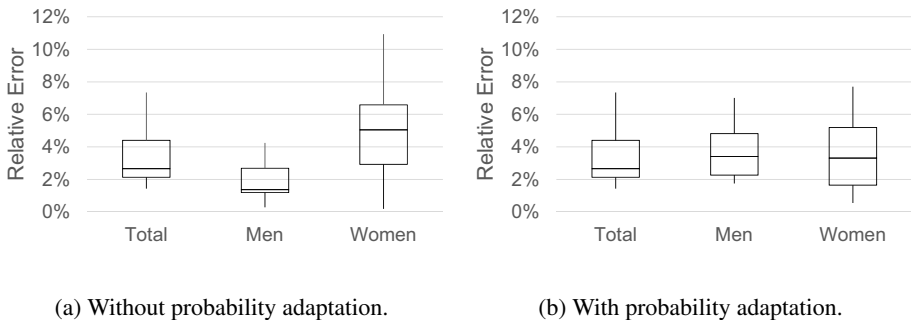


Figure 2. Relative error in Lower Saxony in 2011 with and without probability adaptation.

The total plots in Figures 2a and 2b are identical and reflect the relative error of the number of persons generated between the 100m grid and the municipality level. The aim of the procedure should be to distribute the error as evenly as possible over the individ-

ual characteristics in order to achieve stable convergence. The generation without PA in Figure 2a shows a clearly higher variance, in particular the even distribution of the error on the characteristics does not take place, because a low error for one gender results in a high error for the other gender. Without PA, for example, the maximum error can vary between 4% and 12%, while with PA it can vary between 7% and 8% as Figure 2b indicates. Overall, PA is necessary as a factor for the stabilization and reliability of the procedure, whereas the synthesized population for Lower Saxony deviates from the actual population by an average of 2.7% and can therefore be used for the analysis of local accumulations.

6. Conclusion

Currently, \bar{p} is used for 2011 as an identity function or to represent the population for other years. It is conceivable that this function can also be trained to minimize errors, so that overall errors and thus also the generation error of the characteristics decrease and can lie within the 99% confidence interval. In addition, the procedure can also be used for regions other than Lower Saxony or for data with other granularities. Furthermore, the method can also be applied for the approximation of health information described in [6], e.g. supply chain company networks on different levels (personnel, specialist groups, company groups) for drug research or for clinical studies, in order to dissolve clusters if necessary and to reduce study costs by data approximation in preliminary studies.

This paper has given an overview of various granularities of demographic data in epidemiological cancer registration and a probabilistic procedure for approximating actual demographics for a 1km grid using a combination of Iterative Proportional Fitting (IPF) and cellular automata (CA). For this purpose, municipality-level data were combined with data from the 2011 census for Lower Saxony, so that a statistical distribution was created by IPF as a framework for the actual generation of the population by CA in a 100m grid. The generation error in relation to known data is on average 3%, but a detailed evaluation on 100m or 1km grids was not possible because no reference data were available.

References

- [1] C. Stegmaier, S. Hentschel, F. Hofstädter, A. Katalinic, A. Tillack, and M. Klinkhammer-Schalke, *Das Manual der Krebsregistrierung*. D-82110 Germering/München: W. Zuckschwerdt Verlag GmbH für Medizin u. Naturwissensch., 2018.
- [2] D. Ballas, T. Broomhead, and P. M. Jones, "Spatial microsimulation and agent-based modelling," in *The Practice of Spatial Analysis*, pp. 69–84, Springer, 2019.
- [3] A. Moretti and A. Whitworth, "Evaluations of small area composite estimators based on the iterative proportional fitting algorithm," *Communications in Statistics-Simulation and Computation*, pp. 1–18, 2019.
- [4] R. M. Basse, H. Omrani, O. Charif, P. Gerber, and K. Bódis, "Land use changes modelling using advanced methods: Cellular automata and artificial neural networks. the spatial and explicit representation of land cover dynamics at the cross-border region scale," *Applied Geography*, vol. 53, pp. 160–171, 2014.
- [5] J. W. Weibull, "What have we learned from evolutionary game theory so far?," tech. rep., IUI Working Paper, 1997.
- [6] O. Tamburis and I. Bonacci, "Clusters and communities: raising the bar towards open innovation 2.0 paradigms," *International Journal of Pharmaceutical and Healthcare Marketing*, vol. 13, no. 3, pp. 288–305, 2019.

GenoShare: Supporting Privacy-Informed Decisions for Sharing Individual-Level Genetic Data

Jean Louis RAISARO ^{a,1}, Juan Ramón TRONCOSO-PASTORIZA ^b,
Yamane EL-ZEIN ^c, Mathias HUMBERT ^d, Carmela TRONCOSO ^b,
Jacques FELLAY ^a and Jean-Pierre HUBAUX ^b

^aLausanne University Hospital, Lausanne, Switzerland

^bEPFL, Lausanne, Switzerland

^cCSEM, Neuchâtel, Switzerland

^dArmasuisse, Lausanne, Switzerland

Abstract. One major obstacle to developing precision medicine to its full potential is the privacy concerns related to genomic-data sharing. Even though the academic community has proposed many solutions to protect genomic privacy, these so far have not been adopted in practice, mainly due to their impact on the data utility. We introduce GenoShare, a framework that enables individual citizens to understand and quantify the risks of revealing genome-related privacy-sensitive attributes (e.g., health status, kinship, physical traits) from sharing their genomic data with (potentially untrusted) third parties. GenoShare enables informed decision-making about sharing exact genomic data, by jointly simulating genome-based inference attacks and quantifying the risk stemming from a potential data disclosure.

Keywords. genomic privacy, privacy-conscious tools, risk quantification, inference

1. Introduction

The growing number of health data breaches [1] and the recent revelations about the use of free genealogical databases for law enforcement purposes [2] are increasingly eroding public trust in social media, genealogy websites (e.g., 23andMe and Ancestry.com) and personalized medicine initiatives (e.g., the U.S. All of Us Research Program and the U.K. 100,000 Genomes Project), worldwide. Guaranteeing individuals' genetic privacy is becoming a major challenge for those who want to encourage sharing of medical data for research purposes and better personalized medicine. In response to this demand, the computer security community has made a remarkable effort in providing solutions to secure the storage [3], and the processing of genetic data [4,5,6,7,8]. Yet, the issue of trust goes beyond the provision of secure storage and processing. It is also intimately related to the question of *control* over what data is shared and what information it conveys. In

¹Corresponding Author: Jean Louis Raisaro, Lausanne University Hospital CH-1010, Lausanne, Switzerland; E-mail: Jean.Raisaro@chuv.ch.

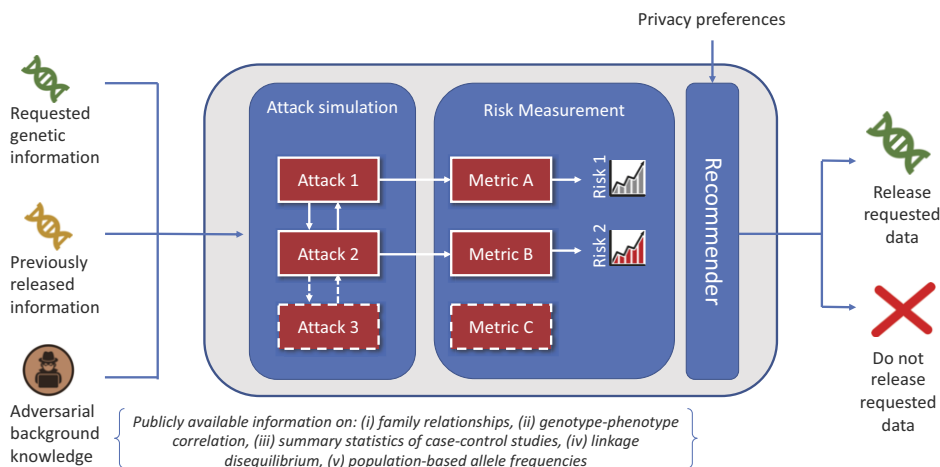


Figure 1. High-level representation of the GenoShare framework.

this paper, we introduce *GenoShare*, the first decision-support framework that provides individuals with guidance at the time of deciding what data can be shared without risking the undesired disclosure of sensitive information.

2. Methods

Let's assume an individual wishes to share part of her genetic profile with a public or private third party in order to join a personalized medicine study or obtain some genetic-related services, but she is concerned about her genetic privacy. Upon reception of a request for genetic data sharing, such an individual can use *GenoShare* to quantify the risk of sensitive attribute disclosure associated to revealing those data. To this end, *GenoShare* simulates the combination of inference attacks relevant to the privacy concerns of the individual (e.g., revealing the participation in a sensitive study, revealing undisclosed kinship, revealing predisposition to certain health conditions) and, based on the information available to the adversary, measures the risk of a potential privacy breach. We consider that a realistic adversary could have access to: (i) part of the individual's genetic profile obtained in the past, (ii) the statistical regularities within the genetic population to which the individual belongs (e.g., linkage disequilibrium, allele frequencies), (iii) the effect-size statistics of the genotype-phenotype association, (iv) the genetic profiles of the individual's relatives who might have already joined the study, and the (v) aggregate statistics from other studies in which the individual might have participated. *GenoShare* is a modular framework and enables different parameterizations (i.e., different levels of adversarial prior knowledge) and the instantiations of different inference attacks, according to the data-sharing context (Fig. 1). Furthermore, *GenoShare* provides a mechanism, based on the generation of synthetic genomes called *avatars*, to protect individuals' genetic privacy from inferences stemming from decision to not release the requested data. Details about the algorithms underpinning *GenoShare* are described in the extended version of this paper [9].

3. Results

To show how GenoShare can, in practice, support privacy-conscious decisions when sharing genetic data, we implement an operational prototype in Python with a Web user interface. We instantiate it with three of the most important genomics-oriented inference attacks that we re-adapted to work on a joint manner and on adversarial partial knowledge: the phenotype inference attack, the membership inference attack and the kinship inference attack. For our experiments, we used individual genetic profiles from the 1,000 Genomes Project [10] and considered different data-sharing use cases where the adversary has access to an increasing amount of information about the targeted individual. For example, for a given individual, GenoShare shows that revealing her genetic variants related to schizophrenia (400 variants) introduces a risk of almost 100% of leaking the value of her predisposition to bipolar disorder and her participation in a study with less than 50 people, and a risk of 60% of leaking her kinship with a first-degree relative who might be known by the adversary. Details and figures about the experiments run to validate GenoShare are described in the extended version of this paper [9].

4. Discussion

Academic solutions for privacy-preserving sharing of genetic data that provide formal guarantees of privacy, such as differential privacy, have mostly focused on data perturbation (e.g., differential privacy). Such solutions, however, have not been adopted by the medical sector as (i) they are only suitable for protecting genetic privacy when aggregate information is shared and (ii) they damage the utility of the data such that is then unusable by practitioners. As opposed to these solutions, GenoShare quantifies the risk of sensitive attribute disclosure when *individual and exact* genetic data is released by using novel meaningful sensitive attribute-oriented metrics and by considering realistic adversaries with limited and parametrizable background knowledge. Finally, to the best of our knowledge, GenoShare is also the first framework to jointly consider relevant attacks in genomic privacy in presence of incomplete information. Therefore, it provides a principled answer to the privacy concerns that affecting the genomic community, and thus it is a firm step forward to enable the responsible and privacy-respecting use of genomic data in research and medical environments.

References

- [1] U.S. Department of Health and Human Services . Breach portal: Notice to the secretary of hhs breach of unsecured [protected health information](https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf). https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf. Last Accessed: October 30, 2019.
- [2] Natalie Ram, Christi J Guerrini, and Amy L McGuire. Genealogy databases and the future of criminal investigation. *Science*, 360(6393):1078–1079, 2018.
- [3] Zhicong Huang, Erman Ayday, Jacques Fellay, Jean-Pierre Hubaux, and Ari Juels. Genoguard: Protecting genomic data against brute-force attacks. In *IEEE Symposium on Security and Privacy*, pages 447–462. IEEE Computer Society, 2015.
- [4] Pierre Baldi, Roberta Baronio, Emiliano De Cristofaro, Paolo Gasti, and Gene Tsudik. Countering GATTACA: Efficient and secure testing of fully-sequenced human genomes. In *ACM Conference on Computer and Communications Security (CCS)*, pages 691–702, 2011.

- [5] George Danezis and Emiliano De Cristofaro. Fast and private genomic testing for disease susceptibility. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*. ACM, 2014.
- [6] Rui Wang, XiaoFeng Wang, Zhou Li, Haixu Tang, Michael K. Reiter, and Zheng Dong. Privacy-preserving genomic computation through program specialization. In *ACM Conference on Computer and Communications Security, (CCS)*, pages 338–347, 2009.
- [7] Paul J McLaren, Jean Louis Raisaro, Manel Aouri, Margalida Rotger, Erman Ayday, István Bartha, Maria B Delgado, Yannick Vallet, Huldrych F Günthard, Matthias Cavassini, et al. Privacy-preserving genomic testing in the clinic: a model using HIV treatment. *Official journal of the American College of Medical Genetics and Genomics*, 2016.
- [8] Muhammad Naveed, Erman Ayday, Ellen W Clayton, Jacques Fellay, Carl A Gunter, Jean-Pierre Hubaux, Bradley A Malin, and XiaoFeng Wang. Privacy in the genomic era. *ACM Computing Surveys (CSUR)*, 48(1):6, 2015.
- [9] Jean Louis Raisaro, Carmela Troncoso, Mathias Humbert, Zoltan Kutalik, Amalio Telenti, and Jean-Pierre Hubaux. Genoshare: Supporting privacy-informed decisions for sharing exact genomic data (extended version). Technical report, EPFL infoscience, 2018.
- [10] The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491:56–65, 2012.

H-Accuracy, an Alternative Metric to Assess Classification Models in Medicine

Andrea CAMPAGNER ^{a,b}, Luca SCONFENZA ^b and Federico CABITZA ^{b,1}

^a *IRCCS Istituto Ortopedico Galeazzi, Milano, Italy*

^b *University of Milano-Bicocca, Milano, Italy*

Abstract. As widely known, regular accuracy is a misleading and shallow indicator of the performance of a predictive model, especially in real-life domains like medicine, where decisions affect health or life. In this paper we present and discuss a new accuracy measure, the H-accuracy, as a more conservative alternative to regular accuracy, which we claim is more informative in the medical domain (and others of similar needs) for the elements it encompasses. In particular, the proposed measure takes into account important information such as the complexity of the cases and the case prevalence in the population. We also provide proof that the H-accuracy is a generalization of the balanced accuracy and illustrate the descriptive power of this score.

Keywords. Accuracy, Machine Learning, Medical Artificial Intelligence, Validation

1. Introduction

Statistical and machine learning communities have developed many metrics to assess and report the predictive performance of a model [1]: some metrics are common and easy-to-comprehend, like accuracy, recall (also known as sensitivity), specificity, the F_1 score (i.e., the harmonic mean of sensitivity and specificity), and the area under the ROC curve (AUROC); others are less common and less straightforward, like the Cohen's Kappa, the Gini coefficient, cross-entropy and Hamming loss.

However, in the non-specialist literature, like medicine, one measure of model performance is much more common than the others: *accuracy*, that is the proportion of correct predictions with respect to the total number of predictions made. Despite its apparent simplicity, accuracy scores can be misleading. For instance, it is generally well known that on imbalanced datasets high values of the accuracy of a model could mirror overfitting with respect to the majority class. In case of datasets where classes are unevenly represented, an alternative measure has been proposed instead of regular accuracy: *balanced accuracy*, that is defined as the average accuracy obtained on each class individually. It can easily be shown that increasing the difference in the prevalence of the two classes of a binary discriminative task makes regular and balanced accuracy diverge pro-

¹Corresponding Author: Federico Cabitza, federico.cabitza@unimib.it

gressively, so that the meaningfulness of the former and more common type of accuracy, which is blind to class prevalence problems, can be disputed¹ [2].

We also believe that regular accuracy is only a partial way to convey meaningful information on the usefulness and practical value of a predictive model, mainly because it overestimates the model performance and does not take into account important characteristics of the test set, like the class (im)balance, the complexity of the cases correctly identified or the importance of rightly detecting one class at the expense of the other(s), according to the users' preferences.

In this paper we present a new accuracy measure, the H-accuracy, or Ha , which we claim is more informative in the medical domain (and other domain with similar requirements) as it covers all of the aspects mentioned above.

2. Method

In the following we deal with a classification task in which the set of classes (or labels) is $C = \{l_1, \dots, l_k\}$; we have a dataset of instances $S = \{x_1, \dots, x_n\}$ sampled from the space of all possible observations X and the real class of instance x is given by a function $c : X \rightarrow C$. We consider, without loss of generality, only *scoring models*: that is, models m defined as $m : X \rightarrow [0, 1]^k$ where, given $m(x) = (l_{1m}(x), \dots, l_{km}(x))$, where $l_{im}(x)$ represents the *confidence* attached to the event $c(x) = l_i$. We also define $\mathbf{m}(x) = \text{argmax}_{l_i} l_{im}(x)$ as the prediction of m . Accuracy of a model m is defined as $a(m) = \frac{1}{|S|} \sum_{x \in S} \mathbb{1}_{\mathbf{m}(x)=c(x)}$, while *Balanced accuracy* is defined as $Ba(m) = \frac{1}{|C|} \sum_{l \in C} \frac{1}{|\{x \in S : c(x)=l\}|}$.

We then introduce the formula for the H-accuracy of a model m as:

$$Ha(m) \Big| \tau, \mathbf{p}, \mathbf{d} = \sum_{l \in C} (\mathbf{p}(l) * \sum_{x \in S : c(x)=l} \frac{\mathbf{d}(x)}{\sum_{c(x)=l} \mathbf{d}(x)} * \sigma_m(\mathbf{l}(x)|\tau)) \quad (1)$$

Where:

1. $\mathbf{p}(l)$ is the *priority* associated to class $l \in C$ (with $\sum_{l \in C} \mathbf{p}(l) = 1$, that is a representation of the relative importance attached to the fact that the model correctly predicts that label;
2. $\mathbf{d}(x) \in [0,1]$ is the complexity of instance x , that is the intrinsic difficulty of making the correct prediction about the case x ;
3. $\mathbf{l}(x)$ is the prediction score for class l on instance x ;
4. $\sigma(\mathbf{l}(x)|\tau)$ is defined as ²

$$\sigma(\mathbf{l}(x)|\tau) = \begin{cases} 0 & \mathbf{l}(x) < \max_{c \in C} \mathbf{c}(x) \\ \frac{\mathbf{l}(x) - \frac{1}{|C|}}{\tau - \frac{1}{|C|}} & \max_{c \in C} \mathbf{c}(x) \leq \mathbf{l}(x) \leq \tau \\ 1 & \mathbf{l}(x) < \tau \end{cases}$$

¹ The same problem affects the area under the ROC curve, another very common performance metrics in medical AI.

² Notice that when $\tau = \frac{1}{|C|}$ the function is defined as $\sigma(\mathbf{l}(x)|\tau) = \begin{cases} 0 & \mathbf{l}(x) < \max_{c \in C} \mathbf{c}(x) \\ 1 & \mathbf{l}(x) \geq \max_{c \in C} \mathbf{c}(x) \end{cases}$

It is easy to prove that Ha is a proper generalization of the balanced accuracy (Ba) and, thus, a proper accuracy measure:

Proposition 1. Let $C = \{l_1, \dots, l_k\}$ $\tau - \frac{1}{|C|}, \forall i. p(l_i) = \frac{1}{|C|}$ and $\forall x \in X. d(x) = k \in [0, 1]$. Then $\forall h. Ba(m) = Ha(m|\tau, p, d)$.

3. Discussion

In this section, we will share some reflections on the different kinds of accuracy scores that Formula 1 allows to produce, and on the methods that can be applied to set the τ , p and $d(x)$ parameters.

3.1. Confidence, for reliability

First, let us consider *confident accuracy* (i.e. $Ha(\tau) = Ha(m|\tau, \frac{1}{|C|}, k)$), according to values of τ , that is the accuracy of a system that is at least τ , *confident* of its output, such that we can claim that “it does not take guesses”.

This measure finds motivation in how human experts behave: in situations of uncertainty they “don’t usually toss a coin” ($\tau = 50\%$) but rather abstain, at least temporarily, and either look for stronger clues or evidence [3], or refer to the second opinion of a more expert colleague [4]. When multiple options are available (i.e., in multi-class tasks), experts can choose the option (e.g., the diagnosis) that is more likely or plausible among the ones available, but they would still act as the option were true only when the plausibility is high enough (otherwise, at least in medicine, a “wait and see” attitude is generally recommended [5]).

Although there are some machine learning applications that apply abstention in conditions of uncertainty [6], this approach is not commonly adopted, and hence it is common that decision support systems have users considering recommendations that are associated with low prediction scores and have a higher chance to be wrong, often without their knowledge, and hence mislead them. Setting a τ , and reporting a corresponding confident accuracy ($Ha(m|\tau)$), can help users evaluate this risk.

Setting a high τ and relying on Ha instead of regular accuracy raises the bar for the performance of machine learning in real-world settings and reflects the requirement to have really accurate models that minimize “guessing” behaviors in a sensitive context.

3.2. Priority, for tailorability

Prioritized accuracy (i.e., $Ha(p) = Ha(m|\frac{1}{|C|}, p, k)$) is defined according to p values. These can be set naively: if $Ha(m|p)$ must express a clear preference for specificity, p could be associated with 0.75 (and 0.25 with respect to sensitivity); and vice versa in case of a preference for sensitivity. Otherwise, if no preference is expressed, a balanced weighting $p = .5$ can be adopted.

On the other hand, priority p can be set empirically in two ways: either explicitly or implicitly. The explicit way entails to collect the preferences of a panel of experts for the classification task at hand by means of a psychometric questionnaire. The implicit case occurs when experts are involved as dataset labelers, and p is set according to their performance.

3.3 Complexity, for usefulness

With the term *complexity*, as hinted above, we refer to decision effort and difficulty to classify an instance for the test set.

But how should $d(x)$ be determined, and hence the *practical accuracy* (i.e. $Ha(d) = Ha(m|\frac{1}{|C|}, \frac{1}{|C|}, d)$) score? Here, we recall that $d(x)$ is a function returning a complexity score for each case x , that is how hard it is to correctly classify the case. This function can be determined in different ways, among which we can distinguish between direct and indirect methods. In regard to the former approach: we can ask the doctors involved to express whether they found a case x difficult or not (0 or 1), as if we asked them if that case deserves to be shared with a class of residents or trainees, or even with the specialist community, as a case report [7]. Similarly, we could ask them to rate how difficult they perceived the case x on an ordinal scale.

Alternatively, once a gold standard is available or has been established, also indirect methods are possible; for instance $d(x)$ can be set on the basis of the error rate of the raters involved for that specific case x : the higher, the more difficult the case is; or on the basis of the confidence the raters express in assessing the case, irrespective of whether their answer was correct or not.

Why assessing complexity is important, and therefore *practical accuracy* is? Think of a very accurate decision support for trivial or easy cases: the medical experts would not consult it, or if they do, would neglect and override it with no regret or doubt in case they disagreed with it³. Conversely, complex cases are those for which even experts would like to receive confirmation or, if lost about them, a proper aid. In these cases, right advice can have a strong impact on the final decision, and wrong advice can mislead the decision makers, a situation that could be worsened by automation bias (i.e., over-reliance on the support, which could be induced by a higher, almost oracular, accuracy exhibited by the system on the easy cases).

For these reasons, we deem the *practical accuracy* an important measure to compute and report, to understand the real utility of a decision support in naturalistic (i.e., real-world) decision making by expert practitioners.

4. Conclusions

In this paper we aimed to raise awareness that the medical community (like many others that increasingly rely on predictive models and decision support systems) must have single, summative but yet comprehensive and informative ways to report and communicate model performance and skills. Other contributions have discussed the limits of common performance measures (e.g., the AUC [8] and accuracy itself [9]) and proposed how to minimize the odds of misinformation about model performance. In addition to this body of literature, we make the claim that the current use of the accuracy indicator for understanding and reporting how well a model works (even assuming a proper testing process producing such an indicator) falls short of communicating the most important feature of a decision support, that is being useful in daily practice.

³Let us just hint at the impact of these "easy misses" on doctors' trust and on the perceived reliability of the support.

Different research groups and companies across the world in the domain of machine-learning based decision support struggle to surpass competitors even by a few percentage points in accuracy (or related measures), with a common “psychological” threshold to interest the medical stakeholders set to 80% at least [10]: besides these figures, in this paper we make the point that is more important to inform stakeholders and practitioners by also showing them alternative and more conservative accuracy measures, with a *clear clinical and practical meaning*.

For this reason, in this paper we proposed a general formula that takes into account the reliability and relevance of the advice given; we also pointed out meaningful areas of the resulting function to focus on specific aspects of the model performance, like reliability (cf. *r*), practicality (cf. *d*) and priority (cf. *p*), and some empirical values to report a set of accuracy measures that could inform the users adopting a model exhibiting such skills, namely *confident*, *prioritized*, and *practical accuracy*.

These measures regard a *static* assessment of the accuracy of a prediction model, evaluated *holistically* on a whole test dataset, as an indicator of the usefulness of the model. Our future work will explore a more dynamic and case-dependent assessment, yielding indicators of the reliability of the model’s predictions: this is the case of a sort of *local accuracy* metrics, that is related to the single new case to classify, evaluated on the basis of the performance of the model on the *n* (e.g., 10) most similar cases in the test set, as a function of their similarity to the new case at hand.

Once more comprehensive static metrics and convenient dynamic metrics have been made available to the users, further research should be aimed at evaluating as the user experience could change and possibly improve when decision makers can learn these additional information about their decision support [11], to mitigate automation bias and increase meaningful use and the overall accuracy of a community of decision makers in real-world settings.

References

- [1] M. Sokolova, G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437, 2009.
- [2] F. J. Valverde-Albacete and C. Peláez-Moreno. 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox. *PLOS ONE*, 9(1):1–10, 01 2014.
- [3] G. A. Lin and R. F. Redberg. Addressing overuse of medical services one decision at a time. *JAMA internal medicine*, 175(7):1092–1093, 2015.
- [4] B. Palmieri and et al. The second opinion medical network. *Int J Pathol and Clin Res*, 3(1):1–7, 2017.
- [5] D. Grady and R. F Redberg. Less is more: how less health care can result in better health. *Archives of internal medicine*, 170(9):749–750, 2010.
- [6] A. Campagner, F. Cabitza, and D. Ciucci. Three-way classification: Ambiguity and abstention in machine learning. In *IJCRS 2019*, 280–294, 2019.
- [7] C. Ortega-Loubon et al.. The importance of writing and publishing case reports during medical training. *Cureus*, 9(12), 2017.
- [8] J. M. Lobo et al. Auc: a misleading measure of the performance of predictive distribution models. *Global ecology and Biogeography*, 17(2):145–151, 2008.
- [9] M. Sokolova et al. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence*, pages 1015–1021. Springer, 2006.
- [10] Alexandru G Floares. Using computational intelligence to develop intelligent clinical decision support systems. In *CIBB: 6th International Meeting on CIBB*, pages 266–275. Springer, 2009.
- [11] E.S. Berner. Diagnostic decision support systems: How to determine the gold standard? *Journal of the American Medical Informatics Association*, 10(6):608–610, 2003.

Heimdall, a Computer Program for Electronic Health Records Data Visualization

Niels MARTIGNENE^{a,1}, Thibaut BALCAEN^b, Guillaume BOUZILLE^c,
Matthieu CALAFIORE^d, Jean-Baptiste BEUSCART^a, Antoine LAMER^a,
Bertrand LEGRAND^d, Grégoire FICHEUR^a, and Emmanuel CHAZARD^a

^a Univ. Lille, CHU Lille, CERIMEA2694, F-59000 Lille, France

^b Univ. Lille, EA2694, F-59000 Lille, CH St Quentin, F-02100 St Quentin, France

^c Univ Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France

^d Univ. Lille, CERIMEA2694, General practitioner, F-59000 Lille, France

Abstract. *Introduction.* Electronic health records (EHR) comprehend structured and unstructured data, that are usually time dependent, enabling the use of timelines. However, it is often difficult to display all data without inducing information overload. In both clinical usual care and medical research, users should be able to quickly find relevant information, with minimal cognitive overhead. Our goal was to devise simple visualization techniques for handling medical data in both contexts. *Methods.* An abstraction layer for structured EHR data was devised after an informal literature review and discussions between authors. The “Heimdall” prototype was developed. Two experts evaluated the tool by answering 5 questions on 24 clinical cases. *Results.* Temporal data was abstracted in three simple types: events, states and measures, with appropriate visual representations for each type. Heimdall can load and display complex heterogeneous structured temporal data in a straightforward way. The main view can display events, states and measures along a shared timeline. Users can summarize data using temporal, hierarchical compression and filters. Default and custom views can be used to work in problem-oriented ways. The evaluation found conclusive results. *Conclusion.* The “Heimdall” prototype provides a comprehensive and efficient graphical interface for EHR data visualization. It is open source, can be used with an R package, and is available at <https://koromix.dev/files/R>.

Keywords. Electronic health records, Visualization, Timeline, Feature extraction.

1. Introduction

Electronic health records data usually contain structured data, such as administrative data, diagnoses (e.g. ICD10 codes [1]), procedures, administered drugs, medical devices, laboratory results, and non-structured data, such as free-text medical records and medical imaging. All of them are characterized by their temporal aspects (e.g. the date of a measurement, the period of a diagnosis, etc.), and relate to one patient. Structured data often take the form of multivalued qualitative variables, encoded using terminologies, and supported by a hierarchical tree of sections and codes.

¹ Corresponding Author: Niels Martignene, Faculté de Médecine, CERIM, F-59035 Lille cedex, France; niels.martignene@protonmail.com.

Data visualization is interesting within 2 contexts. “Transactional” visualization relates to visualization of one patient’s data, in order to provide the patient with appropriate care. “Decisional” visualization relates the visualization of several patient’s data, for management or research purposes [2,3], notably in the frame of data reuse.

In both contexts, the main objective of data visualization is to provide the user with a comprehensive view, and to enhance the mental process of feature extraction [4], by making it easy to detect temporal associations. For instance, in case of INR elevation and red cells decrease, a physician infers the patient encountered a hemorrhage. However, data complexity may overload the user, leading to interpretation errors [5].

Many visual interfaces have been developed for transactional EHR data visualization [6], such as *LifeLines* [7], *MultiMedia Stream System (MMSS) TimeLine* [8], *KNAVE-II* [9], *TimeLine* [10] *HARVEST* [11], or other prototypes [12–14]. However, to our knowledge, such representation methods are rarely implemented in commercial hospital information systems. Some visual interfaces have also been developed for decisional EHR data visualization, such as *OutFlow* [15], *Care Pathway Explorer* [16], *EventFlow* [17], or other prototypes [18–25]. Many of them are dedicated to specific tasks. The objective is to propose a visualization method that could either be used for transactional or decisional visualization of EHR data, and to evaluate it.

2. Method

An abstraction layer for EHR data was first defined, based on an informal literature review, and discussions between authors, taking profit from their experience in clinical care and data reuse. The prototype was intended to be embeddable in C++ programs, R [26], and web pages.

To evaluate the tool, two medical experts (not involved in the development) were asked to retrieve medical information over 24 cases of acute kidney injury (AKI, according to the KDIGO criteria [27]). The 24 cases were represented through 3 interfaces: the classical interface (“classical”: 1 page per type of information, with tables), the manually optimized interface (“manual”: the same, but manually filtered, to present only relevant data), and the Heimdall interface (“Heimdall”: developed in this project, without any prior data filtering). Expert *A* used “classical” to evaluate records #1-8, “manual” for records #9-16, and “Heimdall” for records #17-24. Expert *B* used the same interfaces respectively for records #17-24, then #9-16, then #1-8. Both experts were asked to answer five questions for each case (“d0” relates to the date of AKI, which was intentionally not explicit): (Q1) Was a nephrotoxic treatment discontinued, at d0 or d1? (Q2) Was there a urine dosage of sodium and potassium, at d0 or d1? (Q3) Was there a search for proteinuria, at d0 or d1? (Q4) Was there a urinary tract imaging at d0, d1 or d2? (Q5) Was there an ICD10 code of AKI? Number of errors and time required to answer were compared using an analysis of variance.

3. Results

3.1. EHR data abstraction

We identified 3 types of information: (1) *events*, which occur at a single point in time, and have a label and a date (e.g. a medical procedure, a document), (2) *States*, which are

characterized by a label, a start date and a stop date (e.g. a hospital stay, a diagnosis), and (3) *Measures*, which are characterized by a date, a label and a value (with or without normality range; e.g. a laboratory result, a clinical parameter).

For each of kind of information, the label could be part of a hierarchical tree. In such case, the hierarchy folding had to result in the merging of the symbols. Moreover, additional information could be provided, e.g. the content for a document, which is a sort of *event*, or the unit for a laboratory result, which is a sort of *measure*.

3.2. Prototype development

Heimdall is a read-only visualization tool, developed in C++, using OpenGL, and Dear ImGui. It is available for MS Windows and GNU/Linux. It can be used through its R package, which enables data loading and terminology handling, and can also be integrated inside web pages (through WASM and WebGL).

Pieces of data relating to the same data type are grouped into a component (e.g. 3 values of blood potassium → a unique curve of kalemia). Each component follows a time axis from the left to the right. Components are designed to be stacked vertically.

Figure 1 shows the representation of components according to their type, and their behavior in case of combination (e.g. in case of terminology folding or temporal compression). *Events* are represented as triangles, *states* as transparent rectangles, and *measures* as curves. When folded, measures are transformed into red events in case of abnormal values (outside normality range), or blue events otherwise. The interpolation method can be set as Last Observation Carrier Forward, linear, or cubic spline.



Figure 1. Unfolded and folded representation of 3 types of components

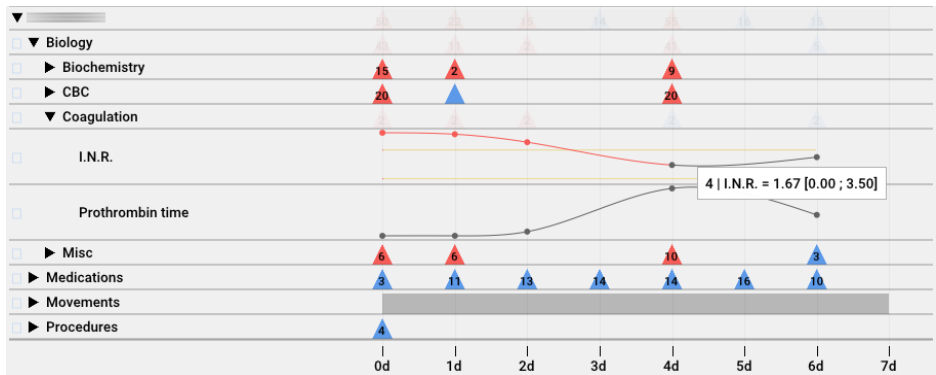


Figure 2. Example of a medical record (partially unfolded view)

Figure 2 illustrates a medical record. The main window principally looks like a stack of components, with a unique time axis. The components are organized according to their hierarchy: (1) patient, then (2) data type (e.g. diagnoses, laboratory results, etc.), then (3) the hierarchical tree of the relevant terminology, if so (e.g. ICD10). This organization is

compatible with both transactional (one patient) and decisional (several patients) usage. In case of multi-patient view, the time axis can be aligned with real dates, or with a chosen event (e.g. birthdate, first inpatient stay, or specific surgical procedure). The representation of big amount of data is made easier by 3 mechanisms: (1) the time axis can be rescaled using control key plus mouse wheel, (2) the top-bottom axis can be compressed, resulting in components folding, and (3) a dropdown menu enables for quick use of problem-oriented filtered views. For instance, the “renal” view only shows diagnoses, procedures, laboratory results and drugs in relation with the kidneys or the renal function. The user can create custom views.

3.3. Prototype evaluation

Loading data from 3,500 patients, including 360,000 values, in Heimdall from R took a couple seconds on a workstation equipped with an old Intel® Core 2 Duo CPU (2009), and required about 50 MB of memory (40% for non-character data, 30% for character data, and 30% for terminologies/graphic interface/polices).

The results of the evaluation are presented in Table 1. For both users, the manually optimized interface and the Heimdall interface were faster to use than the classical interface. The number of errors did not significantly differ.

Table 1. Results of the evaluation (2 experts * 24 clinical cases * 5 questions)

Expert	Measure	“Classical”	“Manual”	“Heimdall”	<i>p val.</i>
#1	Time required (sec, mean & SD)	54.25 (5.42)	35.00 (2.62)	40.88 (5.22)	< 0.001
	Errors (n & %)	0.38 (0.74)	0.12 (0.35)	0.25 (0.46)	0.662
#2	Time required (sec, mean & SD)	57.75 (6.82)	42.25 (5.23)	45.38 (4.47)	< 0.001
	Errors (n & %)	0.38 (0.74)	0.12 (0.35)	0.12 (0.35)	0.546

4. Discussion and conclusion

In this work, we designed a data visualization prototype called “Heimdall”, and evaluated it. The Heimdall interface takes profit from combining foldable hierarchical view with a timeline. Alignment, filtering and custom views enhance data exploration. Heimdall is very fast and easy to use. However, this tool is still not able to display non-temporal data (e.g. birthdate), or data with imprecise timing (e.g. ancient medical history). Non-structured data can be displayed as document icons. Unlike EventFlow [17], Heimdall only provides rudimentary tools for querying and filtering data. The problem-oriented filtering seems to be the most powerful part. Heimdall provides a comprehensive and efficient graphical interface for EHR data visualization. It is open source, and the current R package (alpha version) can be downloaded at <https://koromix.dev/files/R>.

References

[1] World Health Organization, International Statistical Classification of Diseases and Related Health Problems, 2018. <http://www.who.int/classifications/icd/en/> (accessed August 12, 2019).
[2] E. Chazard, and R. Beuscart, Graphical representation of the comprehensive patient flow through the Hospital, AMIA. Annu. Symp. Proc. 2007 (2007) 110–114.
[3] E. Baro, S. Degoul, R. Beuscart, and E. Chazard, Toward a Literature-Driven Definition of Big Data in Healthcare, BioMed Res. Int. 2015 (2015). doi:10.1155/2015/639021.
[4] E. Chazard, G. Ficheur, A. Caron, A. Lamer, J. Labreuche, M. Cuggia, M. Genin, G. Bouzille, and A. Duhamel, Secondary Use of Healthcare Structured Data: The Challenge of Domain-Knowledge Based Extraction of Features, Stud. Health Technol. Inform. 255 (2018) 15–19.

- [5] H. Singh, C. Spitzmueller, N.J. Petersen, M.K. Sawhney, and D.F. Sittig, Information Overload and Missed Test Results in Electronic Health Record–Based Settings, *JAMA Intern. Med.* 173 (2013) 702–704. doi:10.1001/2013.jamainternmed.61.
- [6] R. Bade, S. Schlechtweg, and S. Miksch, Connecting Time-oriented Data and Information to a Coherent Interactive Visualization, in: *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, ACM, New York, NY, USA, 2004; pp. 105–112. doi:10.1145/985692.985706.
- [7] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman, LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records, in: B.B. Bederson, and B. Shneiderman (Eds.), *Craft Inf. Vis.*, Morgan Kaufmann, San Francisco, 2003; pp. 308–312. doi:10.1016/B978-155860915-0/50038-X.
- [8] D.A. Aoyama, J.-T.T. Hsiao, A.F. Cárdenas, and R.K. Pon, TimeLine and visualization of multiple-data sets and the visualization querying challenge, *J. Vis. Lang. Comput.* 18 (2007) 1–21. doi:10.1016/j.jvlc.2005.11.002.
- [9] D. Goren-Bar, Y. Shahar, M. Galperin-Aizenberg, D. Boaz, and G. Tahan, KNAVE II: The Definition and Implementation of an Intelligent Tool for Visualization and Exploration of Time-oriented Clinical Data, in: *Proc. Work. Conf. Adv. Vis. Interfaces*, ACM, New York, NY, USA, 2004; pp. 171–174. doi:10.1145/989863.989889.
- [10] A.A.T. Bui, D.R. Aberle, and H. Kangarloo, TimeLine: Visualizing Integrated Patient Records, *IEEE Trans. Inf. Technol. Biomed.* 11 (2007) 462–473. doi:10.1109/TITB.2006.884365.
- [11] J.S. Hirsch, J.S. Tanenbaum, S. Lipsky Gorman, C. Liu, E. Schmitz, D. Hashorva, A. Ervits, D. Vawdrey, M. Sturm, and N. Elhadad, HARVEST, a longitudinal patient record summarizer, *J. Am. Med. Inform. Assoc. JAMIA.* 22 (2015) 263–274. doi:10.1136/amiajnl-2014-002945.
- [12] W. Aigner, and S. Miksch, CareVis: Integrated visualization of computerized protocols and temporal patient data, *Artif. Intell. Med.* 37 (2006) 203–218. doi:10.1016/j.artmed.2006.04.002.
- [13] FlowingMedia, TimeFlow: Timeline visualization application, 2018. <https://github.com/FlowingMedia/TimeFlow> (accessed April 18, 2018).
- [14] N. Spring, C. Plaisant, G. Marchand, T.D. Wang, V. Mukherjee, D. Roseman, B. Shneiderman, and M. Smith, Temporal Summaries: Supporting Temporal Categorical Searching, Aggregation and Comparison, *IEEE Trans. Vis. Comput. Graph.* 15 (2009) 1049–1056. doi:10.1109/TVCG.2009.187.
- [15] K. Wongsuphasawat, and D. Gotz, Exploring Flow, Factors, and Outcomes of Temporal Event Sequences with the Outflow Visualization, *IEEE Trans. Vis. Comput. Graph.* 18 (2012) 2659–2668. doi:10.1109/TVCG.2012.225.
- [16] A. Perer, F. Wang, and J. Hu, Mining and exploring care pathways from electronic medical records with visual analytics, *J. Biomed. Inform.* 56 (2015) 369–378. doi:10.1016/j.jbi.2015.06.020.
- [17] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman, Temporal Event Sequence Simplification, *IEEE Trans. Vis. Comput. Graph.* 19 (2013) 2227–2236. doi:10.1109/TVCG.2013.200.
- [18] J.L. A, J.L. B, D.N. B, and E.K. A, VizTree: a Tool for Visually Mining and Monitoring Massive Time Series Databases, n.d.
- [19] N.D. Soulakis, M.B. Carson, Y.J. Lee, D.H. Schneider, C.T. Skeeahan, and D.M. Scholtens, Visualizing collaborative electronic health record usage for hospitalized patients with heart failure, *J. Am. Med. Inform. Assoc. JAMIA.* 22 (2015) 299–311. doi:10.1093/jamia/ocu017.
- [20] J. Guerra-Gómez, M.L. Pack, C. Plaisant, and B. Shneiderman, Visualizing Change over Time Using Dynamic Hierarchies: TreeVersity2 and the StemView, *IEEE Trans. Vis. Comput. Graph.* 19 (2013) 2566–2575. doi:10.1109/TVCG.2013.231.
- [21] R. Bellazzi, C. Larizza, P. Magni, and R. Bellazzi, Temporal data mining for the quality assessment of hemodialysis services, *Artif. Intell. Med.* 34 (2005) 25–39. doi:10.1016/j.artmed.2004.07.010.
- [22] K. Wongsuphasawat, J.A. Guerra Gómez, C. Plaisant, T.D. Wang, M. Taieb-Maimon, and B. Shneiderman, LifeFlow: visualizing an overview of event sequences, in: *ACM Press*, 2011; p. 1747. doi:10.1145/1978942.1979196.
- [23] J. Zhao, P. Papapetrou, L. Asker, and H. Boström, Learning from heterogeneous temporal data in electronic health records, *J. Biomed. Inform.* 65 (2017) 105–119. doi:10.1016/j.jbi.2016.11.006.
- [24] M. Vandromme, J. Jacques, J. Taillard, A. Hansske, L. Jourdan, and C. Dhaenens, Extraction and optimization of classification rules for temporal sequences: Application to hospital data, *Knowl.-Based Syst.* 122 (2017) 148–158. doi:10.1016/j.knosys.2017.02.001.
- [25] C. Plaisant, Visualization of temporal patterns in patient record data, *Fundam. Clin. Pharmacol.* 32 (2018) 85–87. doi:10.1111/fcp.12322.
- [26] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2011. <http://www.R-project.org/> (accessed August 12, 2019).
- [27] A. Khwaja, KDIGO clinical practice guidelines for acute kidney injury, *Nephron Clin. Pract.* 120 (2012) c179–184. doi:10.1159/000339789.

How Have Experimental Cancer Interventions Evolved over Time?

Ziran LI^a, Huanyao CHEN^a, Chi YUAN^a, Chunhua WENG^a

^aDepartment of Biomedical Informatics, Columbia University, New York, NY, USA

Abstract. We performed a trends analysis of experimental cancer interventions. The complete records of 32,623 interventional neoplasm clinical trials involving 454 types of neoplasms from 2000 to 2017 were downloaded from the AACT database. The conditions and drug concepts were normalized using MetaMap. The normalized frequencies (NF) for each type of intervention were calculated and compared. Among 95,440 interventions, 77.4% were drugs, 5.3% were radiation, 6.6% were surgery and 10.6% were other therapies. Among 47,754 arms, 82.8% were mono-type interventions and 17.2% were multi-type interventions. Among 73,889 drug interventions, immunologic factor drugs increased rapidly over the last five years. Both breast cancer and pancreatic cancer have been testing new drugs in clinical trials; however, more drugs have been tested in phase 3 or 4 trials and employed in comparator arms for breast cancer compared to pancreatic cancer. Breast cancer trials showed a more even drug NF distribution than pancreatic cancer trials. The JS Distance among three periods (2000-05 vs. 2006-11 vs. 2012-17) showed unidirectional research progress trend for breast cancer, but reverse trend for pancreatic cancer. This study contributes a large-scale landscape overview of the trends in cancer experimental interventions and a methodology for using public clinical trial summaries for understanding the evolving cancer research.

Keywords. Clinical Trials, Trend Analysis, Data Science

1. Introduction

The World Health Organization predicted that worldwide cancer cases are projected to increase by 50% by 2030, from 14 million to 21 million. Well-identified progress in cancer research is pivotal for our understanding of the cancer research direction. Many systematic reviews have uncovered the unbalanced medical advances across different cancers. While prior studies reviewed the evolution of cancer treatment and highlighted the advances in specific cancer types, they are unable to spark more far-reaching insight due to several potential study limitations. *Firstly*, these studies have largely focused on one cancer at a time. *Secondly*, only a qualitative description of the milestones in the cancer treatment history is given based on the reports of domain experts. A longitudinal quantitative analysis is needed to investigate multiple cancer types together to help us better understand the progress in a big picture, compare advances in cancer treatment across different interventions, and guide the future development of cancer therapies. The ClinicalTrials.gov houses all trials including eligibilities, conditions, study design, interventions, result and outcome information. In this study, we propose quantitative methods to answer two questions: (1) What types of interventions have consistent patterns across cancers? (2) What are the trends of intervention research across cancers?

2. Methods

2.1. Dataset

As of July 2018, the record of 271,514 distinct clinical trials were all downloaded from the database of Aggregate Analysis of ClinicalTrials.gov (AACT) [1]. First, interventional studies with primary purpose of treatment conducted within 2000 to 2017 were extracted. Then, an oncology interventional clinical trial dataset was created by selecting clinical trials whose condition MeSH term has the semantic type of “neoplastic process”. The information of treatment arms and interventions were extracted. Each intervention in the clinical trial was recognized as an instance that we focused on.

2.2. Data Preprocessing

Clinical trial conditions were mapped to MeSH concepts. Then all free-text drugs extracted from table of interventions with either “Drug” or “Biological” intervention type were mapped to NCIt concepts using MetaMap. All neoplastic trials were assigned to one of three time periods including 2000-2005 (1st period), 2006-2011 (2nd period), and 2012-2017 (3rd period). The phases of clinical trials were modified as “Phase 1/2” or “Phase 3/4”. Each treatment arm was modified as “experimental” or “comparator” arm type based on the value of group_type field. The interventions were classified into four intervention types, including drug (“Drug” or “Biological”), surgical procedure, radiation, and other therapies (includes “Dietary Supplement”, “Behavioral”, “Device”, “Genetic”, and “Other”). Mapped drug concepts were classified into three drug types based on the UMLS semantic type of the concept, which are immunologic factor (semantic type is “Immunologic Factor”), other bioactive substance (“Hormone”, “Enzyme”, “Vitamin”, “Receptor”, or “Biologically Active Substance”), and conventional drugs (“Pharmacologic Substance” or “Antibiotic”). All interventions in our dataset were stamped with the time period, cancer type, modified phase, modified arm type, modified intervention type, and drug type (only for drug intervention).

2.3. Trend analysis of Intervention Pattern and Drug

We defined the normalized frequency (NF) as the number of interested intervention divided by the total number of interventions for studied condition(s) during a certain time period. The NF of each type of intervention in each year were calculated and compared. We interpreted the change of NF along time as the trends of intervention types. Two types of interventional patterns were defined based on the combination of different types of intervention: mono-pattern includes one type of intervention, such as drug, surgery, radiation or other therapies, while multi-pattern includes the combination of at least two different types of intervention.

We took breast neoplasm and pancreatic neoplasm as examples to demonstrate the different drug research trends across different neoplastic diseases. Breast neoplasm was identified by collecting condition concepts with UMLS CUI including C0006142, C0242788, C0021367, C0206692, C0677776, C0278601, C4018978, C3539878, and C1458155; Pancreatic neoplasm was identified by collecting condition concepts with UMLS CUI including C0030297, C0242363, C1328479, and C0887833. For each trial, we first removed drugs in comparator arms from the drugs in experimental treatment arms to extract research drugs for each trial. The NFs of all research drugs were

calculated in each period. All NFs in each period formed a density distribution of drug tested times in clinical trials. Jensen-Shannon (JS) distance was used to quantify the change of drug distribution among three periods. The density distributions for most popular research drugs were visualized and compared. The number of tested times for most popular research drugs stratified by arm types or trial phase were also visualized and compared between two types of neoplasm.

3. Results

3.1. Clinical trial retrieval and concept recognition

A total of 32,623 neoplastic clinical trials with 47,754 treatment arms encompassing 95,440 interventions were identified, with 73,889 (77.4%) being drugs, 5,048 (5.3%) being radiation, 6,330 (6.6%) being surgery and 10,137 (10.6%) being other therapies. Among 47,754 treatment arms, 39,530 (82.8%) arms were mono-pattern, 8,224 (17.2%) were multi-pattern. The number of newly registered neoplasm clinical trials from 2000 to 2017 displayed an overall increasing trend, corresponding to 454 unique “neoplastic process” concepts and 3439 unique drug concepts. The number of drug intervention was much greater than the number of the other three interventions.

3.2. Interventional Trend analysis

The number of drug interventions far exceeded other types of intervention. It is notable that the ratio among drug, surgery and radiation interventions kept steady since 2000, whereas the number of other therapies had surpassed surgery and radiation after 2006. The multi-pattern (e.g., combining drugs and surgery) has displayed an increasing trend in years after 2010 as opposed to mono-pattern. Figure 1 illustrates the temporal trend of three drug types including immunologic factors, other bioactive substance and conventional drugs. Conventional drugs outnumbered the rest of drug types but exhibited a decreasing trend and the immunologic factor drugs showed an increasing trend.

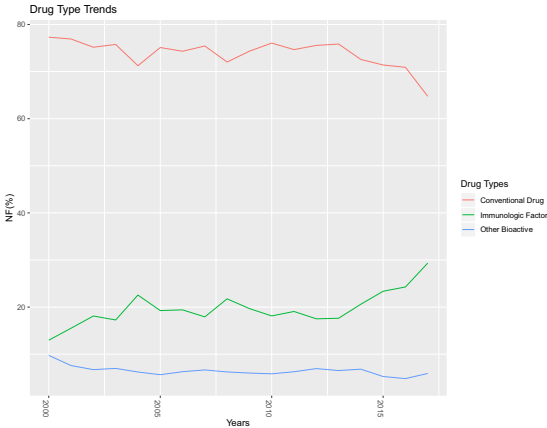


Figure 1. The temporal trend of three types of drug

We took breast neoplasm and pancreatic neoplasm as two examples to compare the drug research trends in different types of neoplastic disease. 3,418 breast neoplasm

clinical trials and 1,271 pancreatic neoplasm clinical trials were extracted from our database. The NFs of all experimental drug were calculated in each period and the result for the top popular research drugs were visualized in Figure 2.

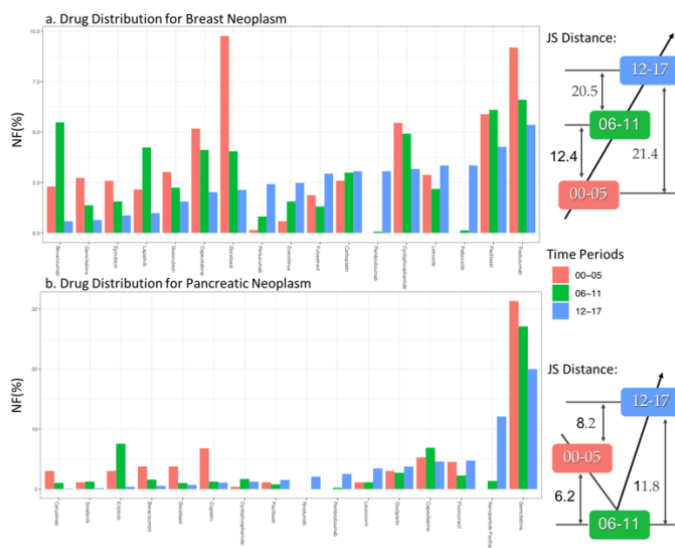


Figure 2. NF distribution of the most popular drugs for (a) breast neoplasm and (b) pancreatic neoplasm.

The drugs were ranked by an increasing order of the NFs in the 3rd period. We observed three different drug NF distributions in three periods for breast neoplasm (Figure 2a). As new drugs studies appeared, the peaks of NF distribution dropped and kept moving rightward to these new drugs from 2000 to 2017. The JS distance of drug NF distributions are 12.4 between the 1st and the 2nd period, 20.5 between the 2nd and the 3rd period, and 21.4 between the 1st and the 3rd period. Thus, the drug research focus showed a unidirectional progress trend for breast neoplasm. In contrast to breast neoplasm, the pancreatic neoplasm has significantly different NF distributions as shown in Figure 2b. However, the NF distributions for pancreatic neoplasm among three periods are pretty similar. Over the past three periods, the drug gemcitabine took the lead with a consistent NF for three periods, and is almost thrice as high as the dominant drugs for breast neoplasm, which indicates the relatively limited drug options for pancreatic neoplasm. The JS distance of NF distributions between the 1st and the 3rd period (8.2) is even closer than the distance between the 2nd and the 3rd period (11.8), which implies a reversing research progress trend for pancreatic neoplasm. For breast neoplasm, all popular research drugs have been tested in phase 3/phase 4 clinical trials to gather more information about the effectiveness by studying different populations with different dosages and by using the drug in combination with other drugs.

4. Discussion

The ratio between drug, surgery and radiation has leveled off since 2000, indicating the popularity of research of three intervention types have stabilized, while drug studies are dominating majority in neoplasm trials. Although drug, surgery and radiation remain the major interventions to treat neoplastic disease, the number of other therapies (supportive

care, dietary supplement, behavioral, observation etc.) tested in clinical trials had surpassed surgery and radiation during recent years. This trend reflects the fact that researchers are paying more attention to supportive care and improving quality of life for cancer patients except for radical cure and prolonging OS [2].

The multi-intervention patterns showed an increasing trend after 2010, which is consistent with the popularity of the concept of multimodality therapy in cancer treatment area. The increasing trend of immunologic factor drugs also implies what current drug research focuses on. This trend can be explained by the fact that immunotherapies generally have fewer side effects than chemotherapies because it boosts the body's natural defenses to fight the disease. Most of immunologic factor drugs are referred as monoclonal antibodies. Among them, we recognize several blockbuster drugs such as Avastin, Rituxna, Herception, and Erbitux. These popular cancer treatments target a wide range of conditions and diseases.

We found that neoplastic diseases whose treatment drugs have steadily progressed, like breast neoplasm, present a more even NF distribution and a unidirectional progress trend between distributions. This unidirectional change implies more new drugs introduced into clinical trials that gradually receive more attentions due to their potential of better efficacy. Neoplastic disease whose drug development is stagnant, such as pancreatic neoplasm, present an uneven distribution and a fluctuating trend, implying researchers encountered difficulty developing more effective substitutes for old drugs.

We further integrated diverse data sources to find evidence supporting our findings. For instance, Bevacizumab, a new drug for breast neoplasm, has a peak frequency in 2006-2011, but its frequency drops drastically within the next period. This result can be attributed to the fact that Bevacizumab was once approved by FDA for metastasis breast neoplasm treatment in 2008 but then revoked in 2011 due to safety concerns including severe high blood pressure and hemorrhaging.

5. Conclusions

This study shows interesting trends in the development of interventions for neoplastic disease in clinical trials and interprets the trends with regard to the corresponding medical contexts and literature. It may generate meaningful insight for the cancer research community so that the medical researchers can quantify the progress of cancer research and utilize prior knowledge of trend. The methodology from this study can also be applied to a diversity of medical fields other than neoplastic diseases.

Acknowledgments: This project is funded by US NLM grant R01LM009886-10.

References

- [1] A. Tasneem, L. Aberle, H. Ananth, S. Chakraborty, K. Chiswell, B.J. McCourt, R. Pietrobon, The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. *PLoS ONE* 2012;7:e33677. doi:10.1371/journal.pone.0033677.
- [2] F. Scotté, The importance of supportive care in optimizing treatment outcomes of patients with advanced prostate cancer. *Oncologist* 2012;17 Suppl 1:23–30. doi:10.1634/theoncologist.2012-S1-23.

Impact of Altering Data Granularity Levels on Predictive Modelling: A Case Study of Fall Risk Prediction in Older Persons

Noman DORMOSH^{a,1}, Ameen ABU-HANNA^a,

Nathalie van der VELDE^b and Martijn SCHUT^a

^aAmsterdam UMC, University of Amsterdam, Department of Medical Informatics,
Amsterdam Public Health Research Institute, Amsterdam, Netherlands

^bAmsterdam UMC, University of Amsterdam, Department of Geriatrics

Abstract. Classification systems are widely used in medicine for knowledge representation. The hierarchical relationships between concepts in a classification system can be exploited in prediction models by looking for the optimal predictive granularity level. In this study, we used the Anatomical Therapeutic Chemical (ATC) classification system to cluster medications in the context of predicting medication-related falls in older persons. We compared the performance of fall risk prediction by describing medications at varying granularity levels of the ATC classification system. We found that the level of abstraction significantly affects the predictive performance in terms of both discrimination (measured by the receiver operating characteristic curve AUC-ROC) and calibration. An implication of these findings to the researchers is that data representation at different granularity levels can influence the predictive performance. The optimal granularity level can be determined by experimentation.

Keywords. classification system, data abstraction, feature selection, prediction, granularity, hierarchy.

1. Introduction

Classification systems describe concepts and relationships between them. They are widely used for knowledge representation. The hierarchical relationship between concepts describes the relationship between a class and its sub-class. For example, the Anatomical Therapeutic Chemical (ATC) classification system [1] describes classes in terms of anatomy, therapy, and chemical compounds of medications and generalization-specialization relationships between medications. Hence a medication can be described at various granularity levels. The choice of the appropriate level of abstraction to describe a class (here medication) depends on the task at hand.

When the task is prediction of an event (such as fall of a person), the problem can be stated as: at which level of granularity should medications be described in order to optimize the prediction of that event. This is not a trivial problem, as the error of a

¹ Corresponding Author. Amsterdam UMC, University of Amsterdam, department of Medical Informatics, Amsterdam Public Health Research Institute, Meibergdreef 9, Amsterdam, The Netherlands; E-mail: n.dormosh@amsterdamumc.nl

prediction model is governed by the bias-variance trade off [2]: using a too coarse level of medications to predict falls may lead to parsimonious models with high bias, while using a too detailed level may lead to high variance. This necessitates search within the hierarchy to find the granularity level that leads to the best prediction model. In essence this is a feature selection strategy in prediction models.

Research studies have linked certain medications with an increased risk of falls in what is called fall-risk-increasing drugs (FRIDs) [3]. FRIDs are grouped in fall risk prediction models to ease interpretation [4, 5]. To exemplify the different granularity levels of these medications, consider “paroxetine”, this medication lies on the leaf level of the ATC hierarchy, and is a sub-class of a group called “selective serotonin reuptake inhibitors”. This group has a super-class named “antidepressants” which belongs to the “psychoanaleptics” class and further to the “nervous system” as the main anatomical level. In studies using medication as predictors, the decision on which level of abstraction to use them has often been arbitrary. In addition, little is known about the impact of medication abstraction on fall risk prediction performance.

The aim of this study was to understand the effect of describing medications at different granularity levels on the performance of fall risk prediction. We achieve this by comparing the predictive performance of models using various granularity levels of the ATC classification system.

2. Materials and methods

2.1. Study approach

In this cross-sectional study, we evaluated six prediction models which included the features pertaining to demographics, comorbidity; and home medications. Medications in the dataset are described at ATC level five (leaf level), which can be grouped at different levels of the ATC classification system. We created six prediction models: ATC-5 a model that uses individual medications; ATC-1 to ATC-4 are models that group medications at the respective ATC level; and one model with only demographic and comorbidity features (No-Med).

2.2. Population, patient inclusion and clinical outcome

Data were obtained from the Medical Information Mart for Intensive Care (MIMIC-III, Version 1.4) [6] database. The MIMIC-III is a publicly available database containing de-identified data from more than 40,000 patients admitted to intensive care settings at the Beth Israel Deaconess Medical Center (BIDMC) in Boston, MA, USA, between 2001 and 2012. The dataset comprises demographic elements such as age, gender and ethnicity, clinical elements including vital signs, laboratory data, diagnostic codes (in International Classification of Diseases 9th Edition, Clinical Modification [ICD-9-CM] format) and outcome values. Beside these it contains various types of clinical notes, including discharge summaries and nursing notes.

We included first admissions of older persons (≥ 65 years of age) with complete discharge summaries.

The clinical outcome was falls before hospital admission. Fallers were identified by the presence of an ICD-9 code that starts with “E88” which represent accidental falls, or

the existence of any of the words (fall, fell, fallen) in the text of the reason of hospital admission field.

2.3. Data extraction and processing

The extracted demographic features were age, gender, ethnicity and marital status. We composed an algorithm to extract home medications and comorbidity features from the discharge summaries. The algorithm utilized RxNorm [7] and the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) in order to match the medications and the comorbidities in the text. Data processing involved the mapping of the extracted medications into ATC codes (level five) using the Observational Health Data Sciences and Informatics (OHDSI) [8] and obtaining all the super (parent) groups for each medication code.

2.4. Statistical analysis

We used the Least Absolute Shrinkage and Selection Operator (LASSO) with logistic regression to allow for feature selection as part of fitting the model and to avoid overfitting [9]. LASSO trades off an increase in bias with a decrease in variance, by adding a penalty against model complexity. This penalty term (λ) is usually chosen by cross-validation. The analysis was performed using the R programming language, version 3.4 and data extraction was performed using Python. We used the implementation in the R *glmnet* package to perform LASSO regression.

We assessed the performance of the models in terms of discrimination (with the area under the receiver operating characteristic curve: AUC-ROC); calibration via calibration plots; and the relative goodness of fit with the Akaike Information Criterion (AIC). To internally validate the results, we applied bootstrapping [10] with 200 recommended repetitions and we calculated the optimism-corrected AUC-ROC. To get an honest assessment and to address LASSO instability of model performance all modelling steps (λ -tuning) were repeated for each bootstrap sample. To test for statistical significance of the differences in the median between each of the models and the model with medications without grouping (ATC-5), we used Wilcoxon signed-rank test. A p -value < 0.05 was considered to indicate statistical significance.

3. Results

The total number of patients was 46,520 in the MIMIC-III database. The number of patients which met the inclusion criteria was 23,034 (49.5%). The number of fallers was 1,721 (7.5% of the included patients). The number of different medications extracted from the discharge summaries was 939.

Figure 1 depicts the relationship between the AUC-ROC and the level of ATC grouping. It can be seen that the AUC-ROC initially increases when climbing up the hierarchy of the ATC, peaks at level two and then drops at the first level of the ATC.

The calibration performance of the models is shown in Figure 2. All the models showed acceptable calibration except the No-Med and ATC-4 models which had miscalibration. As shown in Table 1, the lowest AIC estimate (lower is better) was for the model ATC-2 followed by the model ATC-3. The number of selected features for these two models were 65 and 110 respectively.

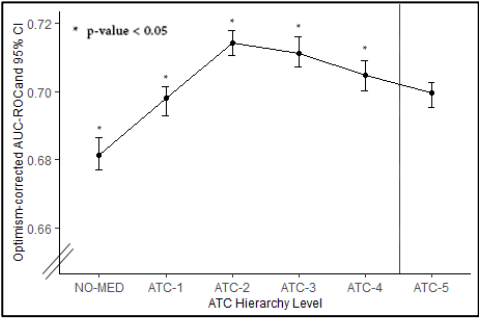


Figure 1. The relationship of AUC-ROC with different levels of grouping using the ATC. Note that the y-axis does not start from zero.

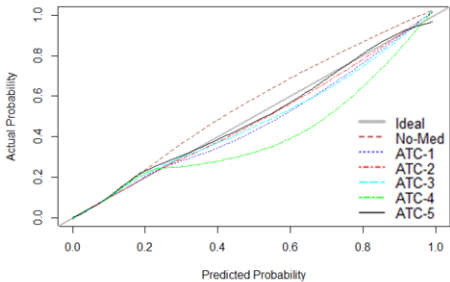


Figure 2. Bootstrap bias-corrected calibration curves of the models.

Table 1. Akaike Information Criterion and the number of selected features after applying LASSO regression.

Models	AIC	#Total Features	#Selected Features
No-Med	-16,406.67	38	28
ATC-1	-16,615.84	52	37
ATC-2	-16,838.66	118	65
ATC-3	-16,655.52	212	110
ATC-4	-16,444.1	465	121
ATC-5	-16,369.3	977	196

4. Discussion and conclusion

This study investigated the effect of leveraging hierarchical relationships of concepts in classification systems for predictive modelling. Our experiments indicate that significant improvement in predictive performance can be achieved by altering granularity levels. This is demonstrated by the ATC-2 model having the highest discrimination performance, acceptable calibration and the lowest AIC.

The strengths of our approach include a systematic evaluation of the performance of the prediction models and the use of bootstrapping to validate the results. The main limitation lies in the fact that the ICD-9 codes were generated for billing purposes at the end of the hospital stay and may not always represent falls before admission. Although the inspection of a random sample of the discharge summaries revealed that most of the patients fell before admission, there still is uncertainty in the definition of fall and the exact moment of its occurrence. This limitation may have affected the overall estimation of falls risk and prediction performance because we considered only home medications

registered on admission and not during hospital stay. Furthermore, we could not validate the accuracy of the algorithms used to extract medications on admissions and comorbidities because of the lack of an annotated dataset. We however, believe these limitations should not qualitatively affect our findings as we evaluated the models under the same conditions.

The utilization of granularity in classification systems was also investigated in other domains. A recent study evaluated the impact of altering bacterial taxonomy on disease prediction [11]. The authors found that describing bacteria at different granularity levels influenced the predictive performance. This finding corroborates what we observed in this work. Another recent study evaluated how different levels of data granularity of medications and diagnoses would affect prediction of diabetic kidney disease (DKD) prediction [12]. The authors showed that different data granularity may not necessarily influence the predictive performance. However, this differs from the findings presented here and this discrepancy could be attributed to the differences in the domains on which the approach was evaluated.

On the basis of this work, we conclude that describing features at different levels of granularity can significantly affect the predictive performance. So far, we considered the same level of granularity to describe medications. A possible interesting extension to our work is to apply tree-lasso [13] or a (greedy-based) search strategy to select a subset of nodes in the tree-structured medication hierarchy. Another important work that warrants further investigation is validating our results on another dataset. Our findings have implications for researchers: we showed that prediction performance can be enhanced by altering the granularity of the data, and the optimal level of granularity can be obtained by experimentation.

References

- [1] Methodology, W.C.C.f.D.S., Guidelines for ATC classification and DDD assignment 2013. Oslo, 2012.
- [2] Friedman, J., T. Hastie, and R. Tibshirani, The elements of statistical learning. Vol. 1. 2001: Springer series in statistics New York.
- [3] de Vries, M., et al., Fall-risk-Increasing drugs: a systematic review and meta-analysis: I. cardiovascular drugs. *Journal of the American Medical Directors Association*, 2018. 19(4): p. 371. e1-371. e9.
- [4] Palumbo, P., et al., Fall risk assessment tools for elderly living in the community: can we do better? *PLoS one*, 2015. 10(12): p. e0146247.
- [5] Homer, M.L., et al., Predicting falls in people aged 65 years and older from insurance claims. *The American journal of medicine*, 2017. 130(6): p. 744. e17-744. e23.
- [6] Johnson, A.E., et al., MIMIC-III, a freely accessible critical care database. *Scientific data*, 2016. 3: p. 160035.
- [7] Nelson, S.J., et al., Normalized names for clinical drugs: RxNorm at 6 years. *Journal of the American Medical Informatics Association*, 2011. 18(4): p. 441-448.
- [8] Hripcsak, G., et al., Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Studies in health technology and informatics*, 2015. 216: p. 574.
- [9] Tibshirani, R., Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996. 58(1): p. 267-288.
- [10] Harrell Jr, F.E., K.L. Lee, and D.B. Mark, Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 1996. 15(4): p. 361-387.
- [11] Bang, S., et al., Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data. *Scientific reports*, 2019. 9(1): p. 10189.
- [12] Song, X., et al., An exploration of ontology-based EMR data abstraction for diabetic kidney disease prediction. *AMIA Summits on Translational Science Proceedings*, 2019. 2019: p. 704.
- [13] Liu, J. and J. Ye. Moreau-Yosida regularization for grouped tree structure learning. in *Advances in neural information processing systems*. 2010.

Incorporation of Multiple Sources into IT - and Data Protection Concepts: Lessons Learned from the FARKOR Project

Doris LINDOERFER^{a,1}, Ulrich MANSMANN^a and Isabel REINHARDT^{a,b}

^a *Institute for medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany*

^b *Medical center of the university of Munich*

Abstract. The IT- and data protection concept of the FAmiliäres Risiko für das KoloRektale Karzinom (FARKOR) project will be presented. FARKOR is a risk adapted screening-project in Bavaria, Germany focusing on young adults with familial colorectal cancer (CRC). For each participant, data from different sources have to be integrated: Treatment records centrally administered by the resident doctors association (KVB), data from health insurance companies (HIC), and patient reported lifestyle data. Patient privacy rights must be observed. Record Linkage is performed by a central independent trust center. Data are decrypted, integrated and analyzed in a secure part of the scientific evaluation center with no connection to the internet (SECSP). The presented concept guarantees participants privacy through different identifiers, separation of responsibilities, data pseudonymization, public-private key encryption of medical data and encrypted data transfer.

Keywords. Data integration, HIC data, pseudonymization, data encryption

1. Introduction

The *FAmiliäres Risiko für das KoloRektale Karzinom* (FARKOR) project assesses the efficacy and safety of a risk adapted screening-program focusing on familial colorectal cancer (CRC) in young adults (aged 25-50 years of age) [1]. It is implemented in Bavaria, Germany [2]. The project integrates individual data from different sources: (1) individual diagnostic and treatment data from medical practitioners, (2) individual patient-reported lifestyle data, and (3) individual outcome data from health insurance companies (HIC).

FARKOR has two parts: (1) Efficient selection of persons with familial colorectal cancer risk within a population of 3 million people, (2) offering safe and efficient screening approaches for the selected high risk population. FARKOR is a prospective, population-based cohort study. Its efficiency is assessed by (1) the number of advanced adenoma or carcinoma detected in the selected high-risk population, (2) a health economic evaluation that estimates avoided morbidity and mortality.

A large number of practices enrolled into the project. All 35 health insurance companies working within Bavaria participate. They provide individual outcome data

¹ Corresponding Author: Doris Lindoerfer, lindoerf@ibe.med.uni-muenchen.

regarding CRC mortality and morbidity. This allows to have control outcome of persons not enrolled in FARKOR.

To implement and to run this project successfully, a reliable IT infrastructure and robust data protection concept has to be established that is also compliant with the new European General Data Protection Regulation (EU-GDPR) [3].

2. Methods

2.1. Several institutions and responsibilities

The FARKOR project integrates data from multiple sources and exchanges the data between different institutions (see Figure 1).

2.2. Specific challenges and used methods

Specific challenges regarding the data protection concept are: 1.) Integrating data from different sources; 2.) Protecting the privacy of participating individuals 3.) Protect business issues between the participating HICs (which patient is insured by which HIC); 4.) Providing the legal reasons for collecting the data (informed consent for subjects enrolled in FARKOR, permission to use individual data of persons not enrolled in FARKOR – the controls); 5.) Performing person rights: Informing about collected data or deleting the records on request; 6.) Providing a 24/7 availability of the IT-infrastructure.

The FARKOR IT- and data protection concept is built on several strategies [4-10]: Informational separation of powers (Informationelle Gewaltenteilung [11]) by using different identifiers for data of different origin [4]; Pseudonymization of identifiers [5] and encryption of personal health information [6]; a trust center [7] and concepts of asymmetric encryption and pseudonymization [8-10].

Integration of HIC data over multiple databases uses standardized terminologies like the International Classification of Diseases, 10th Revision (ICD-10), *Operationen- und Prozedurenschlüssel* (OPS) and *Einheitlicher Bewertungs Maßstab* (EBM).

3. Results

3.1. IT- and data protection concept with data flow

The program started in October 2018. By end of December 2019 about 9000 persons were enrolled. The data capturing, data encryption, data flow (see Figure 1), pseudonymization and data decryption work appropriately.

A specific data service by the resident doctor's association (1) *KVB data service* administrates the individual project data. It is separated from the institution's routine data and general activities. The (2) *trust center* pseudonymizes the data identifiers.

The *KVB data service* provides an internet portal to document the *medical data*: demographic, diagnostic, and treatment data of individuals enrolled in FARKOR. After informed consent, participants are enrolled by their primary care physician who does a basic familial CRC risk assessment. In case of a positive assessment, the participant

will be referred to specialists for further staging and potential screening measures. The *KVB data service* encrypts the data and makes it available to the trust center for download.

Record linkage between the *medical data* and the individual health insurance data uses the social security numbers (KVNR). The trust center downloads the medical data from the *KVB data service*, removes the specific HIC IDs, pseudonymizes the individual IDs, and makes the data available for download to the secure evaluation environment (SECSP). The SECSP is a secure part in the secure evaluation center (SEC) with no connection to the internet.

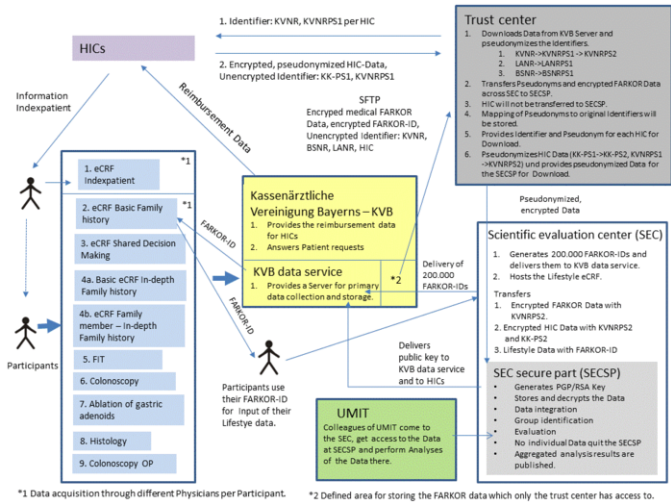


Figure 1. Data flow in the FARKOR project.

The trust center provides at scheduled points in time, the social security numbers (KVNR) and a pseudonym of the KVNR of the FARKOR participants to the HICs for download. The HICs match individual data with related ICD-10, OPS and EBM codes according to the specified identifiers. The HICs pseudonymize the data accordingly, remove the KVNR and deliver the pseudonymized data back to the trust center. The trust center pseudonymizes the HIC data, performs the record linkage and allows to download the data to the SECSP. The HICs data is now doubly pseudonymized.

The concept of informational separation of powers requests, that all medical data is encrypted by the KVB data service and has to be decrypted in the SECSP. Encryption uses public-private-key RSA encryption. Patient-related data are encrypted with the public key. The identifying variables are not encrypted for pseudonymization purposes. It was requested by the LMU's IRB that the trust center cannot read the medial data.

The SECSP downloads the data and decrypts the individual medical data with the private key. Then the SEC performs data storage, data integration and analysis. All data transfers are transport-encrypted.

Furthermore, lifestyle data are important for the analysis. The SEC created individual FARKOR-IDs, which are delivered to the *KVB data service*. These IDs are provided to the participants when they enroll and allows login to the lifestyle questionnaire website. Participants use their internet browser to answer the lifestyle

questionnaire. The corresponding server is hosted in the SEC. These data will be transferred to the SECSP and integrated to the project data with the FARKOR-ID.

The health economic evaluation is carried out by partners at the Private University for Health Sciences, Medical Informatics and Technology (UMIT). They perform their analysis within the SEC. No data leaves the SECSP.

3.2. Warranty of the participants rights

Participants may request at any time a copy of their individual data stored. They can also withdraw their consent of participation in FARKOR at any time. Their requests are handled in a standardized manner. The process for withdrawal is shown in Figure 2. To date, four participants have withdrawn their consent. The corresponding data was deleted from the analysis data. So far, no participant has requested access to their data records.

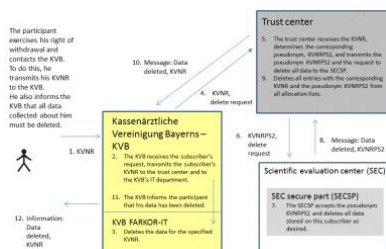


Figure 2. Data flow when participants withdraw their consent.

4. Discussion

The FARKOR IT- and data protection concept enables the integration of multi-source data. To our knowledge, this paper describes for the first time an EU-GDPR conform simultaneous integration of individual HIC as well as project data during the course of the study. The integration of HIC data allows to have control groups of persons not enrolled in FARKOR, as well as data for subjects enrolled that supplements data documented during medical examinations. This makes it necessary to collect and store social security numbers (KVN). This record linkage process is complex. The KVN is stored and pseudonymized in the trust center and communicated to the HICs. The KVN is not transmitted to the SECSP. The medical data will be encrypted in the KVB data service and decrypted in the SECSP. The medical data and the FARKOR-ID are only available in encrypted form to the trust center.

5. Conclusion

To integrate the multi-source data, a strong IT infrastructure as well as an appropriate data protection concept is needed to guarantee participating persons privacy. This is achieved by using different identifiers for data of different origin, pseudonymization of identifiers in a trust center, encrypted data transfer and encryption of patient-related data.

Ethics and approval

The Institutional Review Board (IRB) of the Medical Faculty of the Ludwig-Maximilians Universität München, Munich, Germany has given ethics approval for this study (18-545). The FARKOR data protection concept has been approved by the Bavarian State Data Protection Commissioner. Furthermore, it was approved by the supervisory authorities of the HICs, the Bavarian State Ministry of Health and Nursing and the Federal Insurance Office, this was necessary because also data about persons not participating in FARKOR are provided by the HICs. The FARKOR IT- and data protection concept was also presented and discussed in the (TMF) e.V. [11], working group data protection (AG DS).

Acknowledgements

The FARKOR project is funded by the Innovation Committee of the Federal Joint Committee (G-BA) <https://innovationsfonds.g-ba.de/>: 01NVF17026. DL received funding from FARKOR. IR received funding from DIFUTURE. DIFUTURE is funded by the Bundesministerium für Bildung und Forschung (BMBF): 01ZZ1804B.

References

- [1] KVB, Projekt, FARKOR: Vorsorge bei familiärem Risiko für das kolorektale Karzinom, Available at: <https://www.kvb.de/abrechnung/verguetungsvertraege/farkor/> (Accessed: Jan 7, 2020).
- [2] S. Hoffmann, A. Crispin, D. Lindoerfer, G. Sroczynski, U. Siebert, U. Mansmann, FARKOR: Evaluating the effects of a risk-adapted screening program for familial colorectal cancer in individuals between 25 and 50 years of age in a prospective population-based intervention study, *BMC Gastroenterology*. [submitted].
- [3] General Data Protection Regulation GDPR, Available at: <https://gdpr-info.eu/> (Accessed: Jan 7, 2020).
- [4] D. Lindoerfer, U. Mansmann, IT Infrastructure of an Oncological Trial Where Xenografts Inform Individual Second Line Treatment Decision, *Stud Health Technol Inform*, **235** (2017), 226-230.
- [5] T.M. Deserno, D. Haak, V. Brandenburg, V. Deserno, C. Classen, P. Specht, Integrated Image Data and Medical Record Management for Rare Disease Registries. A General Framework and its Instantiation to the German Calciphylaxis Registry, *J Digit Imaging*, **27:6** (2014), 702-713.
- [6] S. Sherman, O. Shats, E. Fleissner, G. Bascom, K. Yiee, M. Copur, K. Crow, J. Rooney, Z. Mateen, M.A. Ketcham, J. Feng, A. Sherman, M. Gleason, L. Kinarsky, E. Silva-Lopez, J. Edney, E. Reed, A. Berger, K. Cowan, Multicenter breast cancer collaborative registry, *Cancer Inform* **10** (2011), 217–226.
- [7] H. Aamot, C.D. Kohl, D. Richter, P. Knaup-Gregory, Pseudonymization of patient identifiers for translational research. *BMC Med Inform Decis Mak* **13**, 75 (2013).
- [8] T. Michelsen, C. Lins, S. Gudenkauf, A. Hein, C. Lüpkes, Privacy by Design for Integrated Case and Care Management: Receiver-Oriented Encryption in STROKE OWL, *Stud Health Technol Inform*, **258** (2019), 110-114.
- [9] N. Adam, T. White, B. Shafiq, J. Vaidya, X. He, Privacy preserving integration of health care data, *AMIA Annu Symp Proc*, **11** (2007), 1-5.
- [10] M. Bialke, P. Penndorf, T. Wegner, T. Bahls, C. Havemann, J. Piegsa, W. Hoffmann, A workflow-driven approach to integrate generic software modules in a Trusted Third Party, *J Transl Med*. **13**, 176 (2015).
- [11] TMF e.V., Available at: <http://www.tmf-ev.de/> (Accessed: Jan 7, 2020).

Integrating the Comparative Toxicogenomic Database in a Human Pharmacogenomic Resource

François-Élie CALVIER ^{a,1}, Cedric BOUSQUET ^{a,b}

^aPublic Health and Medical Information Unit, University Hospital of Saint Étienne, Saint Étienne, France

^bSorbonne Université, Inserm, université Paris 13, LIMICS, Paris, France

Abstract. Information relevant to pharmacogenomics studies is available in several open databases, which makes it difficult to synthesize the available data. Within the PractikPharma project, several databases were integrated to PGxLOD, a resource dedicated to the generation and verification of pharmacogenomic influence on drug responses. The Comparative Toxicogenomic Database (CTD) describes the toxic effects of many chemicals on living species based on the literature. Since drugs are peculiar chemicals and side effects are peculiar toxic effects, we aimed at extracting information from CTD that matches drug side effects in the human specie.

Keywords. Pharmacogenomics, biomedical genomics, Data interoperability and data integration, Quality of data,

1. Introduction

Pharmacogenomics is a branch of pharmacology that studies how drug responses may be influenced by the genome. As information relevant to discovering and explaining such responses are available in several databases, the consortium of the PractikPharma project implemented methods to extract state of the art knowledge from these databases and the literature. First, PGxO [1], an ontology covering the pharmacogenomic domain was proposed, which focuses on human and drug relative information. Second, this model was implemented in PGxLOD, a pharmacogenomic LOD i.e. a semantic entry (a SPARQL endpoint) to access and query pharmacogenomic data.

The Comparative Toxicogenomic Database (CTD) [2] is a premier public resource for literature-based, manually curated associations between chemicals, gene products, phenotypes, diseases, pathways and environmental exposures. Considering drug as particular chemicals and pharmacogenomic relations as particular toxicogenomic relations, we identified CTD as a potentially interesting pharmacogenomic information resource. In this paper, we describe the method applied to extract data relevant to pharmacogenomic studies from CTD to prepare its integration to PGxLOD.

¹Corresponding Author: François-Élie Calvier, CIM 42, chemin de la marandière, CHU de Saint-étienne Hôpital Nord 42055 SAINT-ÉTIENNE CEDEX 2, France; E-mail: francois.calvier@chu-st-etienne.fr

2. Material and Method

We first describe CTD data. Then, we explain how we selected and extracted data from CTD. Finally, we integrated these data to PGxLOD by mapping elements from the same classes that were previously imported in PGxLOD from other databases.

2.1 Data representation in CTD

CTD provides a collection of biological links documenting toxicogenomic connections. An initial set of links is gathered through a semi-automatic systematic review of pubmed articles. This task consists first, in identifying the elements (i.e. chemicals, genes or diseases) involved in those links. Then, in associating a link with the pubmed IDs of the articles that mention co-occurring elements. Links validated by CTD curators are annotated using controlled vocabularies. In a latter step, these links are integrated with data from external reference databases. For example, chemicals are mapped to the Chemical Abstract Service (CAS) [3], Medical Subjects Headings (MeSH) [4], and DrugBank [5] databases. Genes are mapped to the NCBI gene database [6]. Diseases are mapped to the MeSH, and pathways are mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) [7] and Reactome [8] databases. Finally, additional links are inferred by reasoning over the curated set of links.

Regularly updated, 164,640 chemicals, 470,706 genes, 12,784 diseases, and 2,363 pathways over several species and their known biological links were represented in 2017 (version 14932). Biological links represented in CTD can be either documented knowledge from the literature or inferred knowledge. We detail these links in table 1. When links are inferred, the inference node is specified in parenthesis.

Table 1. Biological links statistics

	Chemicals	Genes	Diseases
Genes	1,482,537	—	—
Diseases	5,245,672	62,519,080	—
Direct evidence	94,065	29,715	
Inferred	5,151,607 (Genes)	62,489,365 (Chemical)	
Pathways	1,119,270	135,816	544,767
Direct evidence	1,119,270	135,816	0
Inferred	0	0	544,767 (Genes)

2.2 Data extraction from CTD

In order to extract human pharmacogenomic data from CTD, drugs should be distinguished from other chemicals, and human genes should be distinguished from genes belonging to others species. We filtered biological links in order to discard those involving non-human genes and non-drug chemicals. Finally, we removed pathways and diseases

without a biological link involving a drug or a human gene. Identification of Drugs was the main difficulty of the extraction process.

Chemicals are described in CTD by their names, and some external databases IDs. These external IDs, in particular DrugBank ID and CAS RN, were exploited to identify drugs among the chemicals. As DrugBank is a drug database, a chemical in CTD with a DrugBank ID is necessarily a drug. The CAS number is a unique identifier defined by the American Chemistry Society (ACS). It is widely used in several databases, and can be found in DrugBank and in KEGG.

2.3 Direct Exploitation

56,536 chemicals in CTD have a non empty CAS number field which is available for 4,789 drugs in DrugBank. This information is also available for 8,760 within the 10,467 drugs present in KEGG. An identification of chemicals which are actually drugs was performed using the CAS number as a pivot.

2.4 Indirect Exploitation

Several databases relative to drugs do not mention the CAS number. Moreover, this information is often incomplete like in DrugBank and KEGG. Therefore, we also exploited the Pharmacogenomics database (PharmGKB) [9] as a reference drug database. KEGG and PharmGKB provide more cross-references than DrugBank (resp. 5 and 19), that can be exploited thanks to web services to reinforce mappings using the CAS number:

- Unichem [10] provides mappings between 38 databases including CAS but was not updated since 2014.
- Chemical Identifier Resolver (CIR) [11] is a NIH web service that includes mappings between CAS, InChI, and chemSpider identifiers.

2.5 Integration of selected data from CTD to PGxLOD

Many databases for life sciences already have been integrated in the Bio2RDF project [12]. Several databases (e.g. DrugBank, KEGG, and PharmGKB) have been integrated to PGxLOD thanks to a mapping between their respective ontologies. Nevertheless, as Bio2RDF has not been updated since 2014, we integrated CTD data to PGxLOD by using a script written by following the Bio2RDF guidelines.

3. Results

3.1 Extraction Results

4,944 drugs were identified within CTD chemicals. DrugBank IDs of 1,710 chemicals are available in CTD which makes identification straightforward for these chemicals. Thanks to the CAS number as a common foreign key 4,363 drugs were identified among CTD chemicals (resp. 1,235 from DrugBank and 3,328 from KEGG). 1,731 of the drugs from KEGG were not identified thanks to DrugBank. Finally, Exploiting web services lead to the identification of 268 additional drugs. From the identified drugs, it is possible to extract biological links that are relevant for human pharmacogenomics. We provide, in

table 2, the number of extracted links with respect to CTD existing links. A gene is involved in a relation with a drug if there is a relation with a chemical from CTD that has been identified as a drug. Thus, 31,559 genes are involved in the 631,143 drug-gene links. Moreover, 19,780 genes of this subset belong to the human genome.

Following the same rule, diseases and pathways relevant to human pharmacogenomics are extracted. The respective numbers are shown in table 3.

Table 2. Extracted Biological links statistics

	CTD	Human pharmacogenomics
Chemical-Gene	1,482,537	631,143
Chemical-Disease	94,065	5,388
Chemical-Pathways	1,119,270	235,430

Table 3. Extracted linked elements

	CTD	Human pharmacogenomics
Chemicals	164,640	4,944
Genes	47,706	19,780
Diseases	12,784	6,796
Pathway	2,363	2,061

3.2 Integrating CTD Data to PGxLOD

Even if all extracted biological links are relevant to human pharmacogenomics, we have chosen to integrate to PGxLOD only those that were manually curated. Since PGxLOD is a semantic resource, all inferred links from CTD could be retrieved by adding an inference rule when using a reasoner. We represented Biological links as objects in order to keep the documentation link information. Thus, a biological link between two biological elements was represented by two properties between each element and the biological link object. Pubmed ID was then added as a property of a biological link in PGxLOD.

4. Discussion

We conducted the extraction of data relevant to pharmacogenomics studies from CTD. This extraction was first expected as an easy task because of the existence of cross references between CTD and DrugBank. But, later, we realised there was a lack of interoperability between data provided by the chemistry and biomedical communities, and that most of the cross references were missing (less than 10% of the DrugBank drug IDs are found in CTD and the CAS RN information is available for less than 50% of DrugBank drugs). Our extraction task is complete with respect to the limitations of our approach.

The integration of CTD in PGxLOD lead to an update of PGxO. Since biological links were relations between two elements and PGxO expects ternary relations, we represented links as an object property in PGxO. CTD enables the possibility to provide the literature link referencing a biological link evidence. To take advantage of this resource, biological links should be represented as a concept with the linked biological elements and the literature references as properties.

Our approach is mainly manual and requires much time to understand chemical and biomedical knowledge. The gap between both communities obstruct our reconciling efforts. Moreover, many database cross-references are out-dating and can lead to uncertain results. Nevertheless, the CTD project is still providing updates and particularly CAS number information for more and more chemicals. As only a third of CTD chemicals were studied in our extraction due to CAS number availability, the process can be

reproduced to improve the extraction with the updated data. Finally, we chose to exploit only invariable information such as IDs to perform an automatic extraction. Information like the name of the chemicals could also be exploited in a semi-automatic process.

Thanks to the integration of CTD to PGxLOD, experiments could be conducted to explore the link between adverse drug reactions and molecular characteristics of the drugs [13].

Acknowledgements

This work is supported by the PractiKPharma project, founded by the French National Research Agency (ANR) under Grant No. ANR-15-CE23-0028. We acknowledge Pierre Monnin who supported the integration to PGxLOD of data we selected from CTD.

References

- [1] Pierre Monnin, Joël Legrand, Graziella Husson, Patrice Ringot, Andon Tchechmedjiev, Clément Jonquet, Amedeo Napoli, and Adrien Coulet. Pgx and pgxlod: a reconciliation of pharmacogenomic knowledge of various provenances, enabling further comparison. *BMC bioinformatics*, 20(4):139, 2019.
- [2] Allan Peter Davis, Cynthia J Grondin, Robin J Johnson, Daniela Sciaky, Benjamin L King, Roy McMorran, Jolene Wieggers, Thomas C Wieggers, and Carolyn J Mattingly. The comparative toxicogenomics database: update 2017. *Nucleic acids research*, 45(D1):D972–D978, 2016.
- [3] HL Morgan. The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. *Journal of Chemical Documentation*, 5(2):107–113, 1965.
- [4] Carolyn E Lipscomb. Medical subject headings (mesh). *Bulletin of the Medical Library Association*, 88(3):265, 2000.
- [5] David S Wishart, Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, and Murtaza Hassanali. Drugbank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*, 36(suppl 1):D901–D906, 2007.
- [6] Donna Maglott, Jim Ostell, Kim D Pruitt, and Tatiana Tatusova. Entrez gene: gene-centered information at ncbi. *Nucleic acids research*, 33(suppl 1):D54–D58, 2005.
- [7] Hiroyuki Ogata, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, and Minoru Kanehisa. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 27(1):29–34, 1999.
- [8] Antonio Fabregat, Konstantinos Sidiropoulos, Phani Garapati, Marc Gillespie, Kerstin Hausmann, Robin Haw, Bijay Jassal, Steven Jupe, Florian Korninger, Sheldon McKay, et al. The reactome pathway knowledgebase. *Nucleic acids research*, 44(D1):D481–D487, 2015.
- [9] TE Klein, M Whirl-Carrillo, EM McDonagh, JM Hebert, L Gong, K Sangkuhl, and CF Thorn. Pharmacogenomics knowledge for personalized medicine. *Clinical pharmacology and therapeutics*, 92(4):414, 2012.
- [10] Jon Chambers, Mark Davies, Anna Gaulton, Anne Hersey, Sameer Velankar, Robert Petryszak, Janna Hastings, Louisa Bellis, Shaun McGlinchey, and John P Overington. Unichem: a unified chemical structure cross-referencing and identifier tracking system. *Journal of cheminformatics*, 5(1):3, 2013.
- [11] Wolf-Dietrich Ihlenfeldt, Johannes H Voigt, Bruno Bienfait, Frank Oellien, and Marc C Nicklaus. Enhanced cactvs browser of the open nci database. *Journal of chemical information and computer sciences*, 42(1):46–57, 2002.
- [12] Alison Callahan, José Cruz-Toledo, Peter Ansell, and Michel Dumontier. Bio2rdf release 2: improved coverage, interoperability and provenance of life science linked data. In *Extended Semantic Web Conference*, pages 200–212. Springer, 2013.
- [13] francois-elie calvier, Pierre Monnin, Miguel Boland, Patryk Jarnot, Emmanuel Bresso, Malika Smail-Tabbone, Adrien Coulet, and Cedric Bousquet. Providing molecular characterization for unexplained adverse drug reactions. In *Medinfo*, 2019.

Integration of Unstructured Data into a Clinical Data Warehouse for Kidney Transplant Screening -Challenges & Solutions

Maximilian ZUBKE^{a,1}, Matthias KATZENSTEINER^a and Oliver J. BOTT^a

^a*University of Applied Sciences Hannover, Hannover, Germany*

Abstract. After kidney transplantation graft rejection must be prevented. Therefore, a multitude of parameters of the patient is observed pre- and postoperatively. To support this process, the Screen Reject research project is developing a data warehouse optimized for kidney rejection diagnostics. In the course of this project it was discovered that important information are only available in form of free texts instead of structured data and can therefore not be processed by standard ETL tools, which is necessary to establish a digital expert system for rejection diagnostics. Due to this reason, data integration has been improved by a combination of methods from natural language processing and methods from image processing. Based on state-of-the-art data warehousing technologies (Microsoft SSIS), a generic data integration tool has been developed. The tool was evaluated by extracting Banff-classification from 218 pathology reports and extracting HLA mismatches from about 1700 PDF files, both written in german language.

Keywords. NLP, image processing, information extraction, data warehouse, graft rejection, kidney transplant

1. Introduction

Due to demographic change an increasing number of severe kidney disease cases and, as a result, an increasing need for kidney transplants can be expected. Early detection of graft rejection is therefore of particular importance in the therapy of kidney transplanted patients.

The joint project "Screen-Reject: A lateral flow test for rejection diagnostics" focuses on innovative diagnostics for the aforementioned purpose and is stratified into three subprojects. The subproject "Screen-Reject: Clinical data warehouse for graft rejection diagnostics" of the project network is concerned with the provision of a clinical data warehouse (CDWH) as a starting point for the development of an expert system to support rejection diagnostics based on clinical data. [1]

In the course of the medical monitoring of patients, both structured treatment-relevant data and unstructured findings texts are recorded. These unstructured findings texts pose special challenges for data extraction.

¹ Corresponding Author, Maximilian Zubke, University of Applied Sciences Hannover, Expo Plaza 12, 30539 Hannover, Germany; E-Mail: Maximilian.zubke@hs-hannover.de.

1.1. Requirements of clinical practitioners and scientists

Our clinical partners are especially interested on the Banff classification [2,3] as well as several information from examinations of human leukocyte antigens (HLA). A requirement analysis resulted in the following four requirements for the CDWH:

- Obtaining the distribution of Banff classification for all the patients included in our study.
- Selecting patients by Banff classification and coding results.
- Estimating the most common HLA mismatches.
- Selecting donor/receiver pairs having certain HLA mismatches.

However, since that information are only described in narrative instead of structured reports, it is not yet possible to select certain variables and filter them by value.

Thus methods are needed to extract the relevant information from the narrative reports.

2. Method

2.1. Data Staging

Initially, reports from the local pathology lab and the clinic for transfusion medicine were acquired. While the pathology reports could be provided as plaintext block from another clinical data warehouse [4], the findings from transfusion medicine were only available as PDF files. Thus the data staging environment have been extended by a data lake. Any information that could not be imported directly into the data warehouse was temporarily stored in the data lake.

2.2. Extracting Banff classification from staged plaintext blocks

The extraction, transformation and loading of defined information from narrative plaintext was realized by a SSIS² package. This package requires a source table containing plaintext resp. file paths to text data and processes the rows sequentially by a JSON-based communication with external text mining functionality. In detail, one or several external text mining tools can be connected with the SSIS package by corresponding script tasks. Finally, the data provided by the script task is written to previous designed tables of the data warehouse. Figure 1 illustrates this workaround.

² SQL Server Integration Services (<https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver15>)

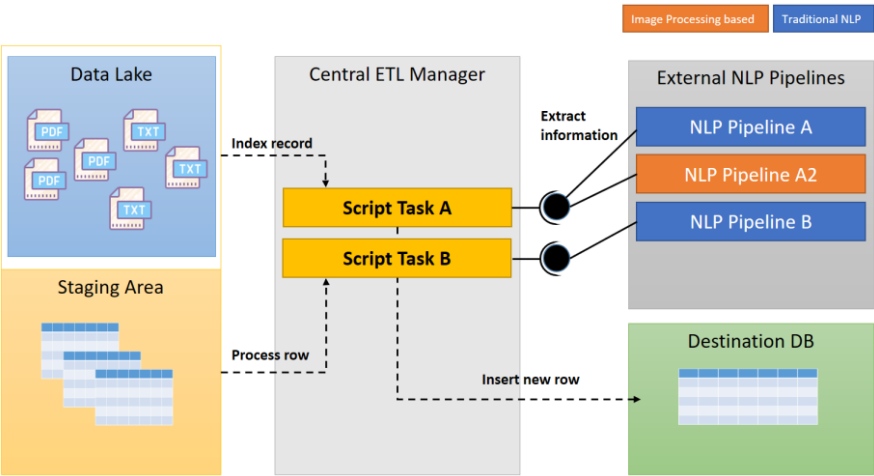


Figure 1. Workflow of the SSIS based ETL package: A central component coordinates the data flow and communicates with external NLP Pipelines via corresponding SSIS Script Tasks.

In the present case, the determined Banff categories [2,3] including optional comments as well as the Banff coding [2,3] should be extracted from pathology reports.

Considering that, two script tasks have been created, getting the same input but interacting with different NLP pipelines (one for each information to be extracted) and sending the output to two different tables in the destination database. The NLP pipelines can be any executable program which generates structured JSON output that is interpretable by the corresponding script task. In our use case, all NLP pipelines were represented by compiled python3 scripts. The information extraction functionality was realized using regular expressions.

2.3. Reverse engineering of HLA mismatch table from unstructured PDF files

In contrast to pathology findings, reports about HLA mismatches could not be made available in form of plaintext blocks, provided by another data warehouse. Instead, the only available descriptions were PDF files generated by third-party software or a scanner.

For this reason, the first step was the transformation of PDF files to text files. A preprocessing step that converts each PDF file to a text file without losing information like headlines or line breaks has been integrated. However, this approach was not able to recognize tables: Visual lines that separate rows and columns were not extracted and thus the whole table was transformed to a text block, whereby internal borders between rows and columns as well as external borders to previous and following document content. Even conversions to Word Documents or other formats failed in most cases. However, from a visual perspective, tables are obvious and clearly delimitable parts of a document. Due to this, we decided to read the table of mismatches by image processing instead of natural language processing. Therefore, the PDF is considered as an image and the first step of this approach is the detection of horizontal and vertical lines [5]. Next, intersections of these lines are calculated. Thereafter, reading the document from left to right, each recognized intersection marks the border between columns and horizontal lines having intersection with the same vertical lines mark the rows.

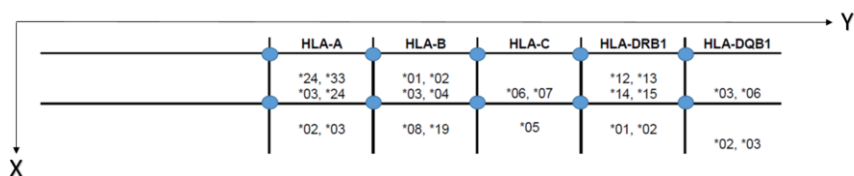


Figure 2. Sample image of a HLA table: Cell values are antigen stereotypes that extend the HLA name in the headline of the same column. Each row describes a sample from a patient.

Finally, the image regions surrounded by the intersection points become extracted and analyzed by an OCR component, OCR errors are prevented by comparing the results with the plaintext of the PDF. Geometric coordinates about the position of lines and intersection points are used to associate regions with each other and deriving the semantic information; For example, in Figure 2 the cell containing “HLA-A” and the cell containing “*02,*03” are surrounded by the same vertical lines and located in the same geometric region on Y-axis. Image regions having the same coordinates on the X-axis are interpreted as fields of the same row.

Overall, using visual features instead of linguistic or keyword based approaches, works fine to recognize the semantic structure of a mismatch table within a PDF-based report.

The extraction of the remaining content of interest, which is represented as a section or list is done by the same approach as described in section 2.2.

3. Result

As illustrated in Figure 3, the requirements mentioned in the introduction were fulfilled by extending ETL-pipelines with methods from natural language processing and image processing.

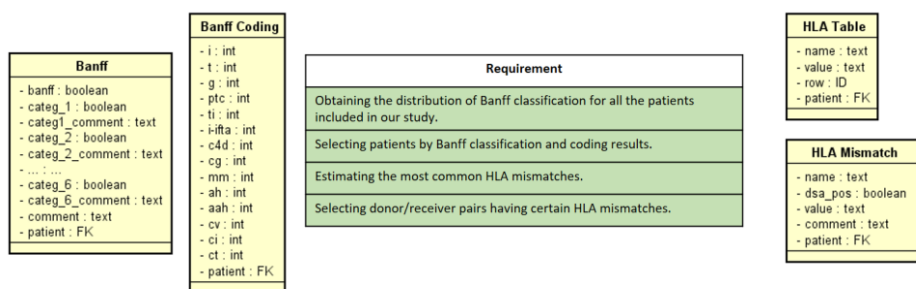


Figure 3. Requirements (green) were fulfilled by providing detailed information tables (yellow).

Technically, the described process was implemented by a generic SSIS package, which can be reused for further, similar data integration projects.

3.1. Evaluation

The described data integration approach was used for 218 pathology reports and 1700 HLA reports, both in German, given by the source dataset provided for this research project.

The correctness of our ETL approach was checked by manual verification of the mapping results. Our approach reaches a precision of 100% and a recall of 100% for pathology reports and 85% for the HLA reports. The latter recall of only 85% has been caused by unexpected document structures. After appropriate modifications, also the NLP Pipeline for the HLA reports reached a recall of 100% and finally all requirements were accomplished.

4. Conclusions

Even this work has confirmed that clinicians tend to document crucial information in form of free texts. Thus, clinical information systems shall be able to extract information of interest from such documentations and transform it into structured representations required for secondary use in research and care. The transformation of information from free text to structured formats is typically done using natural language processing. However, it has been shown, that methods from image processing can outperform traditional NLP in some specific tasks like the extraction of coherent document parts which are marked visually like e.g. tables.

5. Acknowledgement

We want to thank the center for information management of the hanover medical school for providing servers and enabling a connection to their clinical data warehouse.

The research for this paper was done within the joint project “Screen Reject”, founded by the European Regional Development Fund (ERDF) and the State of Lower Saxony, Germany.

References

- [1] Screen Reject: Subproject 3: Clinical data warehouse for NTx rejection diagnostics, <http://screen-reject.f3.hs-hannover.de/about/> [cited 2019 April 11].
- [2] K. Solez, R.B. Colvin, L.C. Racusen, M. Haas et al., Banff 07 classification of renal allograft pathology: updates and future directions, *American journal of transplantation* **8** (2008), 753-760.
- [3] C. Roufosse, N. Simmonds, M. Clahsen-van Groningen, M. Haas et al., A 2018 reference guide to the Banff classification of renal allograft pathology, *Transplantation* **102** (2018), 1795-1814.
- [4] S. Gerbel, H. Laser, N.Schönfeld and T. Rassmann, The Hannover Medical School Enterprise Clinical Research Data Warehouse: 5 Years of Experience, *International Conference on Data Integration in the Life Sciences*, Springer (2018), 182-194.
- [5] R.O. Duda and P.E. Hart, Use of the Hough transformation to detect lines and curves in pictures, *Communications of the ACM* **15** (1971), 11-15.

Interpretation Method for Continuous Glucose Monitoring with Subsequence Time-Series Clustering

Masaki ONO^{1 a}, Takayuki KATSUKI^a Masaki MAKINO^b

Kyoichi HAIDA^c and Atsushi SUZUKI^b

^a IBM Research - Tokyo

^b The Dai-ichi Life Insurance Company, Limited

^c Fujita Health University

Abstract. We propose mini-batch top-n k-medoids to sequential pattern mining to improve CGM interpretation. Medical workers can treat specific patient groups better by understanding the time series variation of blood glucose results. For 10 years, continuous glucose monitoring (CGM) has provided time-series data of blood glucose thanks to the invention of devices with low measurement errors. We conducted two experiments. In the first experiment, we evaluated the proposed method with a manually created dataset and confirmed that the method provides more accurate patterns than other clustering methods. In the second experiment, we applied the proposed method to a CGM dataset consisting of real data from 163 patients. We created two labels based on blood glucose (BG) statistics and found patterns that correlated with a specific label in each case.

Keywords. continuous glucose monitoring, subsequence time-series clustering

1. Introduction

In this paper we propose an interpretation method for extracting characteristic sequential patterns from given pairs of labels and continuous glucose monitoring (CGM) value sequences, facilitating better interpretation of CGM values. The label shows a patient's state such as blood glucose level. The key technology is mini-batch top-n k-medoids that provides accurate cluster centers. By clustering the data, we group similar elements of data to find patterns more effectively.

Device makers provide analysis software for CGM logs². The International Diabetes Center created an analysis schema, the "Ambulatory Glucose Profile," consisting of more than 10 metrics, including mean glucose [1]. This analysis schema handles one patient's CGM logs and does not compare them with other patient's logs. We provide sophisticated scheme for comparing patient's logs.

¹ Corresponding author: e-mail: moonoo@jp.ibm.com

² <http://professional.medtronicdiabetes.com/ipro2-professional-cgm> & <https://www.dexcom.com/clarity>

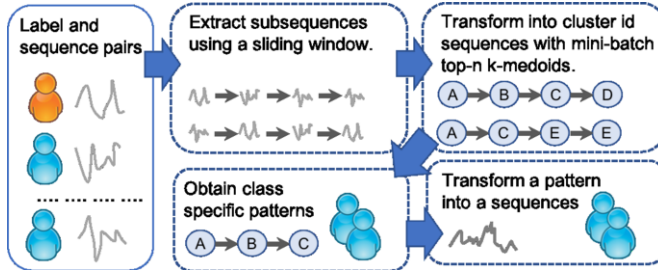


Figure 1. Overview of proposed interpretation method: In this case, the two labels are orange or blue. Our interpretation method extracts the pattern that highly correlates with the blue label.

Medical service workers have to understand patient lifestyles to determine a therapeutic strategy. Thanks to technological advancements in medical devices, we can measure glucose by the minute for weeks using CGM. CGM provides time-series data and has been available since devices with low measurement errors appeared 10 years ago. The review study about CGM [2] reported that there is an urgent need for a standardized interpretation of glucose data and patterns akin to automated electrocardiogram interpretation.

We conducted two experiments, one with a manually created dataset and one with a dataset consisting of actual CGM data. In the first experiment, we evaluated mini-batch top-n k-medoids and determined if our method outperforms other related methods. In the second experiment, we applied our interpretation method to a CGM dataset, that is created through clinical trials, containing 223 waves from 163 real patients over a 2-week period. We extracted interesting patterns from the CGM dataset.

2. Methodology

Our interpretation method consists of the four steps shown in Figure 1. First, we extract subsequences from given sequences by using a sliding window. Next, we apply mini-batch top-n k-medoids to transform CGM value sequences into cluster ID sequences. By clustering the data, we group similar elements of data to reduce noise. Then, we apply sequential pattern mining to extract sequential patterns that correlate with a specific label. Finally, we interpret a label-specific pattern by replacing a cluster ID with a cluster center sequence and obtaining a CGM value sequence.

The procedure of mini-batch top-n k-medoids basically follows the partitional clustering method, but the novel extension uses a mini-batch and n centers in one cluster and runs faster with large-scale data while maintaining clustering accuracy. Some studies have reported that k-means with an Euclidean distance metric for subsequence time-series clustering is useless because the cluster centers are similar to a sine wave [3][4], which is unfavorable to our interpretation task. To address this issue, we use a k-medoid-based method that can use any distance metric and does not create cluster centers based on the Euclidean space.

Let \mathbf{X} denote a set of elements in \mathbf{R}^d and \mathbf{dist} denote a function that returns the distance between two elements. Let \mathbf{C} denote a set of cluster centers and $\mathbf{C}^{(i)}$ denote centers in i -th cluster. Let n denote the number of centers in a cluster, and k denote the number of clusters that is predefined. The method finds a set of cluster centers $\mathbf{C}^{(i)}$ in

$\mathbf{R}^{n \times d}$ that minimize the loss $\emptyset(\mathbf{X}, \mathbf{C}) = \sum_{x \in \mathbf{X}} \min_{i=1}^k \frac{1}{n} \sum_{y \in \mathbf{C}^i} \text{dist}(x, y)$. To minimize the loss \emptyset , we repeatedly update a set of cluster centers \mathbf{C} and each cluster element based on mini-batch $\mathbf{X}' \subseteq \mathbf{X}$ and the following loss functions $\emptyset(x, \mathbf{C}^{(i)}) = \frac{1}{n} \sum_{y \in \mathbf{C}^i} \text{dist}(x, y)$ and $\emptyset(x, \mathbf{X}') = \frac{1}{|\mathbf{X}'|} \sum_{y \in \mathbf{X}'} \text{dist}(x, y)$. When we determine a cluster of $x \in \mathbf{X}'$, we select cluster centers $\mathbf{C}^{(i)}$ that minimizes the loss \emptyset_x , which is the sum of the distance between x and $\mathbf{C}^{(i)}$. When we assign the i -th cluster's centers, we sort the i -th cluster's elements in ascending order based on the loss \emptyset_c , the averaged distance to the elements in the same cluster, and select the top- n elements. At the end of each iteration, we calculate the loss \emptyset to determine whether the loss \emptyset is converged. If the loss is converged, we stop the iteration. Otherwise, we randomly update mini-batch \mathbf{X}' and start the next iteration.

3. Experimental Results

Experiment with the manually created dataset: We created a dataset that contains 500 waves with 16 dimensional attributes. Since the real data has no clear definition for cluster centers, we evaluate our method on the basis of the manually created dataset. First, we created four base waves \mathbf{B} with four dimensions. An element of a wave is a random integer in the half-open interval $[0, 20)$. Second, we created one wave \mathbf{W} by randomly combining the four base waves and multiplying a noise. Let $\mathbf{w}^i \in \mathbf{W}$ and $g \sim \mathcal{N}(\mu = 1.0, \sigma^2 = 0.15)$ be i -th value of the wave and the noise, respectively. Then we define the value with the noise as $\mathbf{w}_i \cdot g$. Finally, by repeating the previous steps, we created a dataset with 500 waves.

We evaluated the clustering performance by measuring the average Earth Mover's Distance (EMD) between the four base waves and the cluster centers. EMD is a distance metric designed to handle time-series data, and lower average EMD indicates a better clustering performance. Let \mathbf{C} be the cluster centers. Let emd be the function that returns the EMD between a pair of waves. The criterion is formulated as $\sum_{b \in \mathbf{B}} \min_{c \in \mathbf{C}} \text{emd}(b, c)$. We applied a clustering method one hundred times with different initial cluster centers and used the average as the clustering performance. We used k-means, k-medoids, DBA [5], and k-shape [6] as baselines. Since k-medoids-based methods can use any distance metric, we applied multiple distance metrics: Manhattan distances (L1), Euclidean distances (L2), EMD, and dynamic time warping (DTW). We set 10 as the number of clusters and 5 as the number of centers in mini-batch top- n k-medoids ($n = 10, k = 5$).

Table 1 shows the evaluation results. We do not list the performance of k-shapes because the performance of k-shapes was over 10 times worse than other clustering methods. We evaluated the performance of mini-batch top- n k-medoids by comparing #2 and #6, #3 and #7, #4 and #8, and #5 and #9. Except for cases with DTW, mini-batch top- n k-medoids performed better than k-medoids. Based on the comparison, the mini-batch top- n k-medoids outperformed the other clustering methods in this evaluation.

Distance metrics defined in Euclidean space were better than EMD and DTW distance metrics designed for time-series data. EMD and DTW handle shift-invariances, one of the biggest issues in time-series data processing. We handle shift-invariances by using a sliding window, not a distance metric. Therefore, EMD and DTW would not interact well with this evaluation.

Table 1. Clustering performance when the number of clusters was 10 (k=10). In this case the clustering performance is the average EMD between cluster centers and four base waves.

ID	Method	Loss	ID	Method	Loss
#1	k-means	8.949	#6	Mini-batch top-n k-medoids + L1	5.185
#2	k-medoids + L1	6.519	#7	Mini-batch top-n k-medoids + L2	5.821
#3	k-medoids + L2	6.000	#8	Mini-batch top-n k-medoids + EMD	7.292
#4	k-medoids + EMD	7.850	#9	Mini-batch top-n k-medoids + DTW	15.061
#5	k-medoids + DTW	14.535	#10	DBA [5]	9.277

Experiment with the real CGM dataset: We applied our interpretation method to a CGM dataset (created through clinical trials) containing 223 waves from 163 real patients over a 2-week period. We targeted patients diagnosed with type 2 diabetes mellitus living in Japan. The target device was Abbott Freestyle Libre³.

We created two labels based on the patients’ blood glucose (BG) statistics. The first label was created from the average of BG and separated the waves into three groups: the average BG is (1) less than 110, (2) more than 110 but less than 140, and (3) more than 140. The numbers of the waves were 129, 80, and 13. The second label was created from the standard deviation of BG and separates the waves into three groups: the standard deviation of BG is (1) less than 30, (2) more than 30 but less than 50, and (3) more than 50. The numbers of waves were 142, 51, and 29. In the experiments, we applied this method by changing the number of clusters, as shown in Tables 2. We set 5 as the number of the centers in mini-batch top-n k-medoids (n = 5) and uses L1 as a distance metric. We used p-value to measure correlation between a label and a pattern. We investigated patterns that occurred more than 100 times and assumed that a pattern whose p-value was less than 0.1 was label-specific.

In the first label, the average of BG, we obtained 686 patterns, and the p-values of 58 patterns were less than 0.1 when there were 30 clusters. When there were 20 clusters, the number of patterns was the largest. In the second label, the standard deviation of BG, we obtained 691 patterns, and the p-values of 58 patterns were less than 0.1 when the number of clusters was 40.

Figure 2 is two examples of label-specific sequential patterns in each label. In both examples, the waves were not observed in group 3, and they highly correlated with groups 1 and 2. When replacing a cluster ID with a cluster representative wave, we selected one that had minimum loss ϕ_x in the cluster. Our aim is to propose an interpretation method rather than give medical insight, so we do not discuss the patterns.

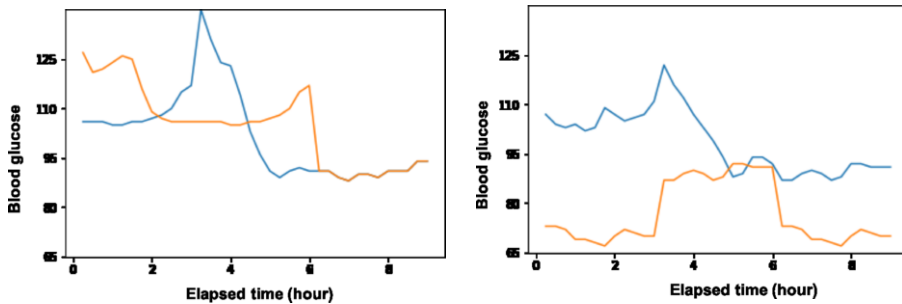
4. Conclusion

In this paper, we focused on a better interpretation of CGM values. We propose an interpretation method of extracting characteristic sequential patterns using mini-batch top-n k-medoids that provides accurate cluster centers. By clustering the data, we group similar elements of data to find patterns more effectively.

³<https://www.freestylelibre.us>

Table 2. Results of the first label created from the BG average (AVG) and the second label created from the standard deviation of BG (STD)

Num. of clusters	20	25	30	35	40	45	50
Num. of patterns (AVG)	768	687	686	680	689	644	659
Num. of label-specific patterns (AVG)	54	50	58	43	48	46	48
Num. of patterns (STD)	791	694	688	691	691	666	676
Num. of label-specific patterns (STD)	53	45	44	43	58	47	56

**Figure 2.** Examples of label-specific sequential patterns produced from the CGM dataset. The waves on the right side were created from the label based on the BG average. The waves on the left side were created from the label based on the standard deviation of BG.

We conducted two evaluations. In the first experiment on manually created dataset, we show that mini-batch top-n k-medoids outperformed the baselines. In the second experiment, we applied our interpretation method to a CGM dataset containing 223 waves created from 163 real patients over a 2-week period. We created two labels based on blood glucose statistics, and extracted label-specific sequential patterns.

References

- [1] R.M. Bergenstal, A.J. Ahmann, T. Bailey, R.W. Beck, J. Bissen, B. Buckingham, L. Deeb, R.H. Dolin, S.K. Garg, R. Goland et al., Recommendations for standardizing glucose reporting and analysis to optimize clinical decision making in diabetes: the Ambulatory Glucose Profile (AGP), Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, 2013.
- [2] D. Rodbard, Continuous glucose monitoring: a review of successes, challenges, and opportunities, *Diabetes technology & therapeutics* **18**(S2) (2016), S2–3.
- [3] J. Lin, E. Keogh and W. Truppel, Clustering of streaming time series is meaningless, in: *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, ACM, 2003, pp. 56–65.
- [4] J.R. Chen, Making subsequence time series clustering meaningful, in: *Data mining, fifth IEEE international conference on*, IEEE, 2005, p. 8–pp.
- [5] F. Petitjean, A. Ketterlin and P. Gançarski, A global averaging method for dynamic time warping, with applications to clustering, *Pattern Recognition* **44**(3) (2011), 678–693.
- [6] J. Paparrizos and L. Gravano, k-shape: Efficient and accurate clustering of time series, in: *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, ACM, 2015, pp. 1855–1870.

Introducing New Measures of Inter- and Intra-Rater Agreement to Assess the Reliability of Medical Ground Truth

Andrea CAMPAGNER ^{a,b} and Federico CABITZA ^{b,1}

^a*IRCCS Istituto Ortopedico Galeazzi, Milano, Italy*

^b*University of Milano-Bicocca, Milano, Italy*

Abstract. In this paper, we present and discuss two new measures of inter- and intra-rater agreement to assess the reliability of the raters, and hence of their labeling, in multi-rater settings, which are common in the production of ground truth for machine learning models. Our proposal is more conservative of other existing agreement measures, as it considers a more articulated notion of agreement by chance, based on an empirical estimation of the precision (or reliability) of the single raters involved. We discuss the measures in light of a realistic annotation tasks that involved 13 expert radiologists in labeling the MRNet dataset.

Keywords. inter-rater agreement, reliability, Ground Truth, Machine Learning

1. Introduction

Data science research needs valid and reliable data to enable the comprehension of the represented phenomena, and sound statistical inference and prediction about those phenomena, that is the definition of predictive models, e.g., by machine learning methods that can help human decision makers classify and interpret the reality of interest.

In many domains, like medicine, phenomena are increasingly measured by sensors, but are still to a large extent described by human observers, who are supposed to produce data in the endeavor to provide accurate and complete representations, by which to detect effects, trends and differences and build reliable prediction models out of them, what is then called *Ground Truth*.

In the light of the unavoidable fallibility of human observers in providing an ever true representation, in domains characterized by high uncertainty, ambiguity and variability of relevant conditions, like medicine, ground truth datasets are built by combining multiple observations and ratings (i.e., labels), and averaging between them, to associate “the one best” label to each case or object of interest. This is the process by which data scientists can bring together subjective ratings (that is how a single rater sees and interprets a given phenomenon) and create a reliable inter-subjective labelling, which is intended to be the most objective representation of the reality of interest [1].

¹Corresponding Author: Federico Cabitza, federico.cabitza@unimib.it

Assessing observer variability, that is the extent multiple raters agree (or disagree) in providing a unique interpretation (that is label, or evaluation), is then important to assess the reliability of the ground truth by which to build predictive models from empirical observations. In the literature there are many measures (e.g., Fleiss' Kappa, Cohen's Kappa, Krippendorff's α) and any of them present pros and cons [2]. In particular, to assess inter-rater agreement, the Krippendorff's α is particularly indicated (although seldom used in medicine and by ML scholars) as it is robust with respect to chance effects and missing values.

However, this measure, as the other ones, adopts a naive model of chance, by which to assess the degree by which multiple observers agree with each others beyond the extent they do so by chance. In particular, no measure considers the expertise of the raters involved, nor their confidence in their specific ratings: in short the reliability of their ratings. This brings us to consider an aspect that is seldom considered: how to assess the reliability of the raters involved, beyond self-assessment? This can be evaluated in many ways: we investigated the relationship between this construct and the extent raters agree with themselves in judging the same phenomenon multiple times over time, and hence they do not take guesses. We call this degree, *self-agreement*, while others refer to it with the expression *intra-rater agreement* (e.g., [3]).

In this paper, we will present two new metrics to assess both self-agreement and inter-rater agreement in order to contribute to the existing literature and provide a better tool to assess the reliability of multi-rater labeled datasets.

2. Method

In this Section we first provide a decision-theory based derivation of a measure for *self-agreement*, intended as the probability that a decision-maker gives the same interpretation of the same phenomenon consistently. Next, we use the self-agreement indicator to define a novel chance-adjusted measure of *inter-rater reliability* that we denote as p . Let $C = \{0, 1, \dots, n-1\}$ be the set of possible class labels, $p = \langle p(0), \dots, p(n-1) \rangle$ be the proportion of 0s, 1s, ..., $n-1$ s in the actual labelings by multiple raters (we assume that these reflect the real class proportions in the reality of interest).

We assume a two-step decision procedure: first the decision maker flips a biased $n+1$ faced coin to choose between *random choice* (x) and *peaked choice* (y_0, \dots, y_{n-1}). Then:

- If the decision maker chose random choice, then she selects one class according to distribution p ;
- Otherwise, if she chose peaked choice, she reports one value (which depends on the specific peaked distribution chosen) with probability 1. Notice that while each peaked distribution assigns probability 1 to a single alternative, the specific peaked distribution is chosen according to (y_0, \dots, y_{n-1}) , which allows to encode possible degree of uncertainty of the rater when she does not guess completely at random.

This decision-making model conforms to standard decision-theory where a decision maker is usually modeled as a probabilistic device in particular, our model represents a generalization of the decision model proposed by Krippendorff as a justification for his α metrics [4], in which raters possess different and unrelated probabilities of selecting among the alternatives at random (e.g. due to their different expertise levels).

How can we find x and y_0, \dots, y_{n-1} given m o_1, \dots, o_m observations of the decision-maker repeating the same annotation task (i.e. annotating multiple times the same case)? Let $d(0), \dots, d(n-1)$ be the observed proportion of 0, 1, ..., $n-1$ labels, respectively, among o_1, \dots, o_m (notice that this is a maximum likelihood estimation of the parameters of a nominal distribution). We can find x, y_0, \dots, y_{n-1} selecting the solution to the following system by maximizing the entropy:

$$\begin{cases} \forall i. d(i) = x * p(i) + y_i \\ x + \sum_i y_i = 1 \\ \forall i. x, y_i \geq 0 \end{cases} \quad (1)$$

We define the *self-agreement* of the observer on label class i as $SA_i = [x * p(i) + y_i]^2 = d(i)^2$ and their overall self-agreement as $SA = \sum_i SA_i$. Notice that SA_i is the probability that the observer gives the same label i when it is asked to give a label to the same object twice and thus it coincides with the *Gini impurity*.

Starting from our definition of self-agreement, we want to define a measure of inter-rater reliability, i.e. a measure of the extent a set of raters agree in labeling a set of cases or objects. Specifically, we want to define a measure ρ that takes into account the fact that, because that raters do not have perfect self-agreements, mutual agreements may arise due to chance; thus, we want ρ to properly discount this case.

Given two raters i, j and a case x we can find the probability

$$P((i, j) \text{ genuinely agree on } x | (i, j) \text{ agree on } x) \quad (2)$$

by a direct application of *Bayes' rule*. Denoting as $y_k^i(x)$ the value of y_k for rater i on case x , we note that this represents the probability that rater i *genuinely* asserts label i on case x . Denoting as $c_i(x)$ the labeling provided by rater i on case x and supposing that $c_i(x) = c_j(x)$, we can see that, assuming that the raters i, j annotate case x independently:

$$P((i, j) \text{ genuinely agree on } x) = y_{c_i(x)}^i(x) * y_{c_j(x)}^j(x) \quad (3)$$

and

$$P((i, j) \text{ agree on } x) = P((i, j) \text{ genuinely agree on } x) + P((i, j) \text{ agree by chance on } x) = y_{c_i(x)}^i * y_{c_j(x)}^j + p[c_i(x)] * x^i * y_{c_j(x)}^j + p[c_j(x)] * y_{c_i(x)}^i * x^j + p[c_i(x)] * p[c_j(x)] * x^i * x^j \quad (4)$$

Then, denoting the ratio of the probabilities in Equations 3, 4 as *Genuine Agreement Effect* (GAE), we can express the ρ measure as:

$$\rho = \frac{1}{|U|} \sum_x \frac{1}{\binom{|D|}{2}} \sum_{i \neq j \in D} \delta(c_i(x), c_j(x)) * GAE(i, j) \quad (5)$$

where $\delta(x, y) = \begin{cases} 1 & x = y \\ 0 & \text{otherwise} \end{cases}$, U is the set of cases and D the set of raters. Notice

that when the sample size to estimate the values of the y_k^i factors is small, the estimate

obtained via the exact methods is not robust when computing the value of ρ , especially if the value were computed on only a small set of cases and then propagated to the remaining ones (e.g. by averaging). In these cases a more robust but approximate estimate can be obtained by considering ignoring the specific probability of getting a particular peaked choice distribution and only considering $\sum_k y_k^i$ as a measure of the ability of rater i of providing self-consistent annotations.

3. Results

In order to test our proposed measures we considered a dataset obtained by a realistic experiment of knee *Magnetic Resonance Imaging* (MRI) annotation. Specifically, we asked 13 radiologists from the IRCCS Orthopedic Institute Galeazzi of Milan (Italy) to annotate 417 MRIs from the well-known MRNet dataset²: in particular, the doctors were asked to assess, for each of these images, the presence of abnormalities (thus, the considered problem was a binary classification setting, i.e. $C = \{0, 1\}$). Furthermore, the doctors were also asked to assess the complexity of each case and the confidence in their annotations, on a 4- and 5-value scale, respectively.

In order to apply our measure, we first had to estimate the self-agreement of each of the raters. In order to do so, we first chose two cases randomly among the ones considered of medium to high difficulty (between 2 and 3 in a 1 to 4 scale) by one of the most experienced radiologist in the panel of experts involved. Then, we inserted these two cases multiple times in the dataset of cases to be labelled, replicating them for 6 times each: in so doing, the dataset had 10 more cases. These replicated images were placed randomly in an annotation sequence (encompassing more than 200 cases) to make more difficult for the radiologists to understand that they had already examined and assessed those cases. Considering the spread, in terms of interquartile range (IQR), in the complexity rating and the varying confidence in the interpretation of these identical cases, we can conjecture that the replicated cases were likely considered different cases by all of the radiologists (First case: complexity IQR = 0.63, confidence IQR = 0.5; Second group: complexity IQR = 0.5, confidence IQR = 0.4).

In regard to the first case, we obtained an average self-agreement of 0.58 ± 0.03 (95% confidence interval, $\min = 0.5$, $\max = 0.72$) For the second group of images, we obtained an average self-agreement of 0.51 ± 0.01 ($\min = 0.50$, $\max = 0.56$): notice that these values are close to the minimal value for SA which, for the binary case, is 0.50. Averaging between the two groups the average self-agreement was 0.54 ± 0.02 ($\min = 0.5$, $\max = 0.64$).

In the computation of the ρ , for each agent i , we considered the values $y_0^i + y_1^i$ estimated on the whole group of 12 repeated images. We obtained a value of $\rho = 0.46$, while for the same dataset we obtained a value of Fleiss' $k = 0.63$ and Krippendorff's $\alpha = 0.63$.

4. Discussion

The above results show how our measure is much more conservative than the others proposed to assess inter-rater agreement. The small difference between the latter measures

²<https://stanfordmlgroup.github.io/competitions/mrnet/>

can be explained by the fact that there was a near perfect class balance: indeed both k and α consider the class balance in order to model chance effects. Furthermore, the large difference between the ρ and the values of the other metrics can be explained observing that the observed self-agreement were indeed very low and the same holds for the obtained y_k^i value (in most cases the value of x^i was near to 0.5). In fact, to obtain a value of $\rho \sim 0.63$ on the dataset, an average GAE around 0.75 would be needed. This shows that, differently from α (and k), the ρ takes into account a model of the rating reliability of the raters involved, and hence it yields a more realistic measure of their agreement: for example, if all raters exhibited a perfect self-agreement (thus $GAE = 1$) we would obtain a value $\rho = 0.82$ while the values for α and k would not change.

Moreover, one could wonder what threshold should be set to assess whether the agreement is sufficient to consider the data reliable, that is what the so-called *smallest acceptable reliability* is [4]. Unfortunately, any proposal of such a threshold would be laden with some extent of arbitrariness. Krippendorff suggests to “not accept data with reliabilities whose confidence interval reaches below the smallest acceptable reliability [...], for example, of .800, but no less than .667” [4](p. 242). An α or an ρ below 0.667 would mean that only two thirds of the data are labelled to a degree better than chance”. This recommendation challenges a much more popular way to interpret agreement scores since the 1970s by [5], that is much more indulgent (a score of .21 is considered an indicator of *fair* agreement; .41 *moderate*; .61 *substantial*).

As said above, differently from α , the ρ takes into account a model of the rating reliability, and hence it yields a more realistic measure of their agreement. In this first formulation, we have considered self-agreement as a proxy of the rater reliability. Since measuring self-agreement with surreptitious repetitions can be intricate, Formula 5 can alternatively be integrated with a measure of the raters’ confidence in the interpretations given (the higher the confidence on a rating, the higher the reliability of that rating as the rater is stating they are not taking a guess). Although this sounds reasonable, yet we did not find confidence to be highly correlated with self-agreement (*correlation* = 0.22, *p* – value = 0.46).

In this regard, however, we should emphasize that low ρ or *SA* scores should not be used as proxy of rating skills, or to judge how good the raters are: in the MRNet study, the 13 radiologists achieved a remarkable average performance of .82 (min: .78, max: .86) in re-annotating the original, low-res dataset on standard monitors and with no incentives. Rather, what low ρ or *SA* scores in ground truthing by experts indicate is the intrinsic ambiguity and complexity of medical phenomena; the over-ambition to pinpoint them with clear-cut labels; and the reckless risk to delegate judgment or advice to classifying machines that interpolate those labels as a way to resolve uncertainty.

References

- [1] L. Mari. A quest for the definition of measurement. *Measurement*, 46(8):2889–2895, 2013.
- [2] A. F. Hayes and K. Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1):77–89, 2007.
- [3] M. E Gianinazzi and et al. Intra-rater and inter-rater reliability of a medical record abstraction study on transition of care after childhood cancer. *PloS one*, 10(5):e0124290, 2015.
- [4] K. Krippendorff. *Content analysis: An introduction to its methodology*. Sage publications, 2018.
- [5] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

Learning Scan Paths of Eye Movement in Autism Spectrum Disorder

Meenakshi CHATTERJEE^{a,1}, Nikolay V. MANYAKOV^b, Abigail BANGERTER^a,
Dzmitry A. KALIUKHOVICH^b, Shyla JAGANNATHA^a, Seth NESS^a and
Gahan PANDINA^a

^aJanssen Research & Development, LLC, USA

^bJanssen Research & Development, LLC, Belgium

Abstract. Eye tracking studies have demonstrated deficits in attention in individuals with Autism Spectrum Disorder (ASD) for a range of different social attention-based tasks. Here we examined social attention skills in a large sample of ASD participants (n = 120), using eye tracking data from a social information processing task, and compared them with a typically developing (TD) group (n = 35). Assuming eye movement parameters are random variables generated by an underlying stochastic process, we modeled the fixation sequences of participants in ASD and TD groups with a Hidden Markov Model. The Regions of Interests (ROIs), modeled as hidden states, corresponded to the true ROIs with a prediction accuracy of >90% for each group. The transition between ROIs revealed bias towards a specific area in the scene in ASD group, which deviated from the TD group. Objective time-dynamic measures of gaze patterns can potentially serve as useful endpoints in ASD diagnosis. **Clinical Trial Registration:** NCT02299700.

Keywords. Autism Spectrum Disorder; eye tracking; Hidden Markov Model.

1. Introduction

Tracking of eye movements allows collection of high-density spatiotemporal data that can provide a rich source of information related to underlying cognitive processes. Several studies have proposed different computational methods to infer observers' characteristics during tasks such as counting, searching, memorizing, or viewing static artificial or natural scenes, that are informative of cognitive processes. These methods broadly range from extracting eye tracking parameters like fixation duration, number, dispersion etc. directly from the data [1]; incorporating spatial information by creating heatmaps, which are 3-D objects (x-coordinate, y-coordinate, fixation density) representing spatial distribution of eye position at a given time [2]; converting eye movement trajectory to a chronological sequence of fixations within pre-defined regions of interest followed by using geometric-based distances metrics to compare scan paths [3]; or using probabilistic approaches to model eye movement [4].

Eye tracking data has been studied as an indicator of social attention in Autism Spectrum Disorder (ASD). Attention to social information is impaired in individuals with ASD, including limited attention to eyes and faces, increased attention to non-social stimuli, increased attention to detail and difficulty with shifting attention compared to typically developing (TD) controls [5]. Commonly, Regions of Interests (ROI) are identified on static images or sequentially on dynamic videos and time spent viewing a prespecified ROI is used to study gaze differences between ASD and TD

¹ Corresponding Author, Meenakshi Chatterjee, Digital Phenotyping, Janssen Research & Development, LLC, Spring House, PA 19477 USA; E-mail: mchatte4@its.jnj.com

groups. Although simple and intuitive, this approach does not capture the dynamic nature of gaze and involves a labor-intensive process of manual labeling of ROIs.

Using data obtained from an observational study collected from the Janssen Autism Knowledge Engine® (JAKE®) [6,7], here we present findings on measurement of eye tracking data from a *Social Information Processing task*. We propose a Hidden Markov Model (HMM) based approach, a time-series probabilistic model, to analyze complex gaze trajectories in ASD and TD groups. Since eye fixation during a task is conditioned on previous fixations, eye movements can be represented as a Markov process. HMM is a natural fit for this problem - it models the specific ROIs participants look at as *hidden states* (eliminating the need to manually label ROIs) and captures the transition among those ROIs (incorporating the temporal dynamics of eye movement).

2. Study design

2.1 Participants: This study was a part of large prospective, 10-week long observational clinical trial conducted at 9 study sites in the US and enrolled 144 ASD and 41 TD participants in total aged between 6 and 64 years [7]. At the study sites, participants completed multiple passive viewing tasks presented on a 23-in. computer screen (1920-by-1024 pixels). Table 1 shows the characteristics of participants for which valid data was available when performing the social information processing task.

Table 1. Characteristics of ASD and TD participants.

	Age (years): Mean (SD)	Age (years): Median (Range)	Male: n (%)
ASD (n = 120)	14.63 (8.25)	12 (6 - 54)	92 (76.67)
TD (n = 35)	16.06 (12.6)	12 (6 - 63)	24 (68.57)

2.2 Social information processing task: Here we focus specifically on the results from a viewing task called the social information processing task. Participants viewed a video of a female actor engaged in child-directed speech against a background of objects, such as toys and an iPad. Detailed description of the video can be found in [8]. The length of the video was 73 seconds with 4 predefined conditions presented over multiple episodes, designed to reflect important aspects of social interaction requiring different demand characteristics: a) Activity: where the actor is making a snack. No direct actor’s gaze or speech cues is present; b) Dyadic Bid: where the actor faces and speaks directly to the camera; c) Joint Attention: where a walking toy robot suddenly begins to move towards the actor, which distracts her from her activity and makes her turn toward that toy; d) Moving Toy: where the toy robot is moving, but the actor looks away at a different toy placed on a wooden stand, located diametrically opposite to the moving toy robot. Figure 1 shows snapshots from the 4 conditions.

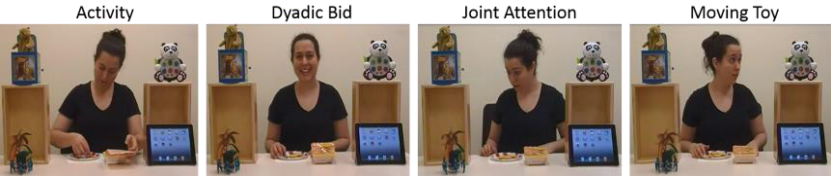


Figure 1. Representative snapshots from 4 predefined conditions in social information processing task.

Participants freely viewed the video while sitting in a comfortable chair ~60 cm from the screen. Tobii X2 eye tracker, mounted below the screen, collected participants’ eye movement data at a sampling rate of 30 Hz. iMotions Biometric Research Platform was

used for stimulus presentation, synchronization and automatic calibration. Before start of the task, a cartoon was placed on a center of the screen to cue participant's attention.

3. Analyses

3.1 HMM for eye movement analysis: A Markov process is a memory-less stochastic process where the probability distribution of the next event or state (in a sequence of events) depends only on the current state and not on the sequence of states preceding it. An HMM is based on augmentation of a Markov process where the system to be modeled has unobservable or *hidden* states. HMMs are suitable for modeling sequential data varying over time and is specified by the following components:

$S = s_1 s_2 \dots s_K$	a set of K states
$A = a_{11} \dots a_{ij} \dots a_{KK}$	a transition probability matrix A , a_{ij} being the probability of moving from state \hat{i} to state \hat{j} , s.t. $\sum_{j=1}^K a_{ij} = 1 \forall i$
$O = o_1 o_2 \dots o_T$	a sequence of T observations
$B = b_i(o_t)$	emission probabilities, each representing the probability of an observation o_t being generated from a state \hat{i}
$\pi = \pi_1 \pi_2 \dots \pi_K$	Prior/initial probability distribution of states, s.t. $\sum_{i=1}^K \pi_i = 1$

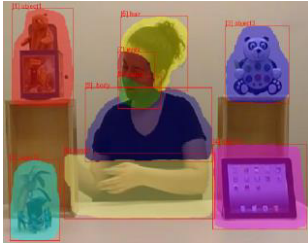


Figure 2. ROIs manually drawn on areas of scene.

In the context of eye movement modeling, the fixation sequence of a participant constitutes a sequence of observation for the HMM. The hidden states of HMM represent Regions of Interest (ROIs) when the participant is viewing the video. The possible ROIs that participants could be viewing during the video is shown in Figure 2. Here we assumed that the distribution of eye fixations on each ROI follows a Gaussian distribution, so the emission densities corresponding to the fixations distribution in each ROI was modeled as two-dimensional Gaussian distribution:

$$B = N_i(m_i, \Sigma_i)_{i \in [s_1, \dots, s_K]} \quad m_i \text{ is the center, } \Sigma_i \text{ is the covariance of } i^{th} \text{ state emission.}$$

We modeled the three ROIs: *hair*, *eye* and *mouth* as a single hidden state and called it *face*. The transition from one ROI (hidden state) to another represents a saccade, and the probabilities of these transitions are summarized in the transition matrix. The prior specifies the probability distribution of the first fixation. The fixation sequences of all participants in each group (ASD and TD) were modeled using a single HMM. To avoid convergence to local minimum, the ASD (and TD) model was trained 100 times with different initial parameters for Gaussian distributions of ROIs, and the model with the highest log-likelihood of the data was selected. The resultant model parameters obtained were the transition matrix, emission densities and priors. From the transition matrix, we quantified the randomness associated with transitioning among different ROIs, by computing the entropy of trajectories of a Markov process as derived in [9].

3.2 Validation of predicted hidden states with Regions of Interest (ROI): To assess if the hidden states or the ROIs predicted by the model for each group correspond to the true ROIs (Figure 2), normalized confusion matrices were generated. Confusion matrix visualizes the performance of HMM model where the rows and columns represent the instances in true and predicted ROIs respectively.

4. Results

Figure 3 shows the accuracy of predicted ROIs when validated with true ROIs for the ASD group (Figure 3A) and the TD group (Figure 3B). Overall prediction accuracy for all ROIs taken together were 92.73% and 94.55% for ASD and TD groups respectively.

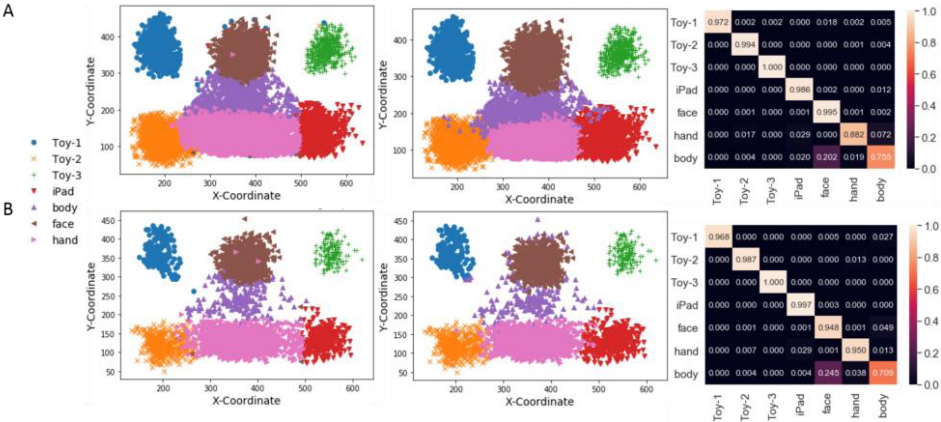


Figure 3. (A) True ROIs (left), predicted ROIs (middle), normalized confusion matrix (right) for ASD group. (B) True ROIs (left), predicted ROIs (middle), normalized confusion matrix (right) for TD group.

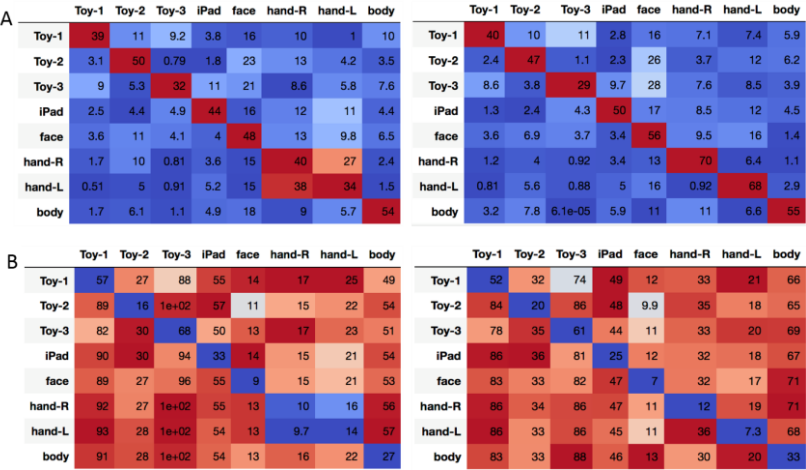


Figure 4. (A) Transition matrices for ASD (left) and TD (right) group; each cell contains probability values multiplied by 100; (B) Markov entropy matrices for ASD (left) and TD (right) group.

All 3 toys in the video array, *iPad*, *face* and *hands* were predicted with high accuracy (>85%) in each group, however accuracy for prediction of *body* fell to 75.5% and 70.9% for ASD and TD groups respectively. The model automatically splits the *hands* into 2 regions of interest: the left hand, *hand-L* and the right hand, *hand-R* (for the purpose of an easy visualization, *hand-L* and *hand-R* are combined and represented as *hands* only in Figure 3). The probability distribution of first fixation indicated that *body* is the most probable initial fixation both for the ASD group (prior probability = 0.95) and TD group (prior probability = 1), with *iPad* being the next most likely initial fixation for ASD group. The fact that both groups initially fixate on *body* could possibly be a manifestation of the experimental design, wherein participants' gaze could still be directed at the cartoon on center of the screen (shown at the beginning of

the task) which falls within the *body* ROI, even slightly after task has started. Figure 4A shows the transition matrix for each group. The TD group exhibited a higher probability to continue fixating at the *face* when they are already fixated at the *face*. ASD group also deviated from the TD group in the way they scanned the *hands*- when participants in the ASD group fixate at the right hand, they are likely to next shift their gaze to the left hand. Similarly, when ASD participants fixate at the left hand, they are likely to next shift their gaze to the right hand. Thus, ASD group was more likely to switch between the two hands than TD group. This stereotypic nature of scanning the *hands* was also captured by the entropies of the transition matrices (Figure 4B), where the entropies of switching between two *hands* (as well as entropies of switching to *hands* from any other ROI) were much lower in ASD group than that in TD group.

5. Conclusion & Future Work

Here we presented a simple methodological framework based on Markov chain and information theory to cut complex eye tracking trajectories into interpretable elementary measures with the least amount of data manipulation or introduction of free parameters as possible. This approach identified the ROI observers look at without requiring manual pre-specification and captured the dynamic nature of gaze patterns by modeling the transition probabilities of the ROIs. One HMM was built for each group, hence could not capture the variability in viewing patterns across participants which can be relevant particularly in ASD group. The ASD group also spent significant time viewing the stimulus background which could bias the choice of hidden states of the model. Finally, the primary focus of the main study was to maximize collection of data in the ASD group, which led to a smaller and unbalanced TD group. Future work will augment the model by introducing a prior to HMM that will model spatial distribution of the moving ROIs, incorporating time spent viewing non-stimulus, building an HMM for each participant and correlating the participant level measures with clinical scales.

Acknowledgement: The authors thank all study participants, Dr. Bennett Leventhal, Dr. Robert Hendren, Dr. Matthew Goodwin, Dr. Frederick Shic, Dr. Geraldine Dawson, Andrew Skalkin and Matthew Boice. This study was funded by Janssen R&D, USA.

References

- [1] Kardan O, Berman MG, et al. (2015). Classifying mental states from eye movements during scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.*, 41(6), 1502.
- [2] Lao J, Miellet S, et al. (2017). iMap4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling. *Behav Res Methods*, 49(2), 559.
- [3] Manyakov NV, Bangarter A, et al. (2018). Visual exploration in autism spectrum disorder: exploring age differences and dynamic features using recurrence quantification analysis. *Autism Res*, 11(11), 1554.
- [4] Zhang J, Chan AB, et al. (2018). Individuals with insomnia misrecognize angry faces as fearful faces while missing the eyes: an eye-tracking study. *Sleep*, 42(2), zsy220.
- [5] Frazier TW, Strauss M, et al. (2017). A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. *J Am Acad Child Adolesc Psychiatry*, 56(7), 546.
- [6] Bangarter A, Manyakov NV, et al. (2019). Caregiver Daily Reporting of Symptoms in Autism Spectrum Disorder: Observational Study Using Web and Mobile Apps. *JMIR Ment. Health*, 6(3), e11365.
- [7] Ness SL, Bangarter A, et al. (2019). An observational study with the Janssen autism knowledge engine (JAKE®) in individuals with autism spectrum disorder. *Front Neurosci*, 13, 111.
- [8] Plesa SD, Brukilacchio B, et al. (2019). Do minimally verbal and verbally fluent individuals with autism spectrum disorder differ in their viewing patterns of dynamic social scenes? *Autism*, 1362361319845563.
- [9] Ekroot L, Cover TM. (1993). The entropy of Markov trajectories. *IEEE Trans. Inf. Theory*, 39(4), 1418.

Leveraging PubMed to Create a Specialty-Based Sense Inventory for Spanish Acronym Resolution

Alexandra POMARES-QUIMBAYA ^{a,1}, Pilar LÓPEZ-ÚBEDA ^b, Michel OLEYNIK ^c
and Stefan SCHULZ ^c

^a Pontificia Universidad Javeriana, Bogotá, Colombia

^b Universidad de Jaén, Jaén, Andalucía, Spain

^c Institute for Medical Informatics, Statistics and Documentation,
Medical University of Graz, Austria

Abstract. Acronyms frequently occur in clinical text, which makes their identification, disambiguation and resolution an important task in clinical natural language processing. This paper contributes to acronym resolution in Spanish through the creation of a set of sense inventories organized by clinical specialty containing acronyms, their expansions, and corpus-driven features. The new acronym resource is composed of 51 clinical specialties with 3,603 acronyms in total, from which we identified 228 language independent acronyms and 391 language dependent expansions. We further analyzed the sense inventory across specialties and present novel insights of acronym usage in biomedical Spanish texts.

Keywords. Acronym resolution, Clinical specialties, Natural Language Processing, PubMed, Sense inventory, Spanish

1. Introduction

An important normalization task in clinical Natural Language Processing (NLP) is the identification and resolution of short forms, e.g. “Human Immunodeficiency Virus (HIV)”. Existing approaches can be distinguished between those that focus on documents where the expansion is explicitly contained (e.g. in scientific articles) and those where the resolution of the acronym is left to the reader (e.g. in clinical narratives). While the former case is mostly addressed by rules, heuristics, and statistical analyses, the latter is usually addressed by using external dictionaries. In both cases, ambiguity arises when an identified acronym may have different expansions (e.g. RA for *Right Atrium* or *Rheumatoid Arthritis*). The problem of ambiguity has been addressed by analyzing the contexts in which the acronym may exist; however, acronym dictionaries in which expansions are linked with possible contexts rarely exist [1].

Several studies have proposed sense inventories that are automatically obtained from MEDLINE or other openly accessible corpora by exploring acronym-definition

¹Corresponding Author: Pontificia Universidad Javeriana, Cra. 7 No 40-62, Bogotá, Colombia; E-mail: pomares@javeriana.edu.co

pairs commonly found in scientific texts. Some of the strategies considered for English abstracts include manually-created pattern matching rules [2,3]; statistics-based word collocations[4,5]; or a mix of both [6]. A seminal study on this topic has been pre-sented by Schwartz and Hearst [7] who proposed a simple algorithm in which possible candidates identified using length heuristics are used to map parenthesized short form-definition-pairs with an alignment-based strategy.

Languages other than English came into focus for short form expansion more recently. In Spanish, the 2018 BARR2 Task [8] explored clinical acronym disambiguation, with a focus on acronym identification on the first subtrack. The winning team of Sánchez-León [9] proposed the application of regular expressions on content of biomedical websites, while Montalvo et al. [10] applied the Schwartz & Hearst algorithm on SNOMED CT to biomedical websites.

This paper contributes to acronym resolution in Spanish clinical texts through the semi-automatic creation of acronym repositories from non-clinical texts using a resource that guides disambiguation. This resource is composed of a set of acronym inventories classified by clinical specialty and characterized by specialty-specific acronym expansions. Our main goal is to demonstrate that the generation of a sense inventory classified by specialty helps acronym resolution by reducing sense ambiguity. To the best of our knowledge, we are the first to apply the Schwartz & Hearst algorithm to the full Spanish subset of MEDLINE and provide our results stratified by clinical specialty.

2. Method

Figure 1 depicts the construction of the sense inventory classified by clinical specialty.

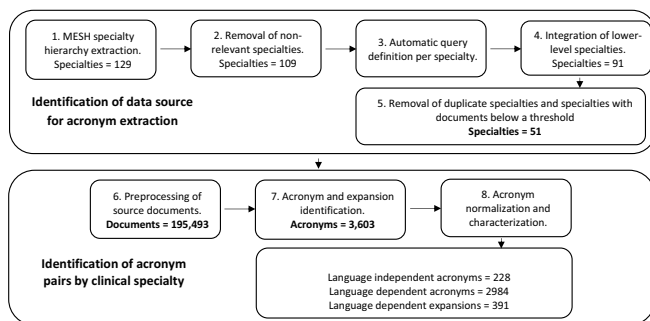


Figure 1. Overview of acronym identification and classification

Identification of data source for acronym extraction: We firstly identified specific branches of the MeSH thesaurus² related to each clinical specialty (step 1). We considered MeSH terms denoting clinical specialties (step 2) and other terms extracted from the “See Also” section of the MeSH hierarchy (step 3) and, from them, only the specialties in the first two levels of the MeSH hierarchy (step 4). Finally, we excluded duplicated

²<https://meshb.nlm.nih.gov>

specialties (e.g. gynecology as “Specialties, Surgical” and as “Reproductive Medicine”) and specialties with less than 250 Spanish titles in their MEDLINE record (step 5).

We explored the MeSH thesaurus and automatically queried documents in PubMed using Biopython³ for matches on either (a) affiliation, (b) title, or (c) MeSH tags.

Identification of acronym pairs by clinical specialty: Once we obtained the relevant documents, we pre-processed their titles and abstracts (step 6) and applied the Schwartz & Hearst algorithm (step 7).

We normalized all acronym expansions to lower case and characterized each acronym/expansion pair by the following features (step 8): (i) intra-frequency, the frequency of the acronym within a clinical specialty; (ii) inter-frequency, the frequency of the acronym across all specialties. (iii) uniqueness across specialties, the number of specialties where the acronym was found; (iv) intra-specialty and (v) inter-specialty ambiguity, the number of different senses found for the same acronym within a specialty or across all specialties, respectively.

We finally (step 9) analyzed the level of language-independence of acronyms and their expansions because we observed acronyms that derived from a different language (mostly English and Latin) together with Spanish expansions. In these articles where acronyms and expansions were identified in both languages, we performed a language independence analysis and classified acronyms automatically into three groups: (i) language-independent acronyms, acronyms and their corresponding expansion are used indistinctly in any language, e.g., “NEWSQOL” (“Newcastle Stroke-specific Quality Of Life measure”); (ii) language-dependent acronym, acronyms whose value and expansion are different for each language, e.g., “MS” (“Multiple Sclerosis”, English) and its corresponding “EM” (“Esclerosis Múltiple”, Spanish); and (iii) language-dependent expansion of acronyms used indistinctly in any language with expansions translated into the document language (the translation does not necessarily coincide with the acronym creation rules), e.g., “ANCA” (“Anti-Neutrophil Cytoplasmic Antibodies”) and “ANCA” (“anticitoplasma de neutrófilos”).

3. Results

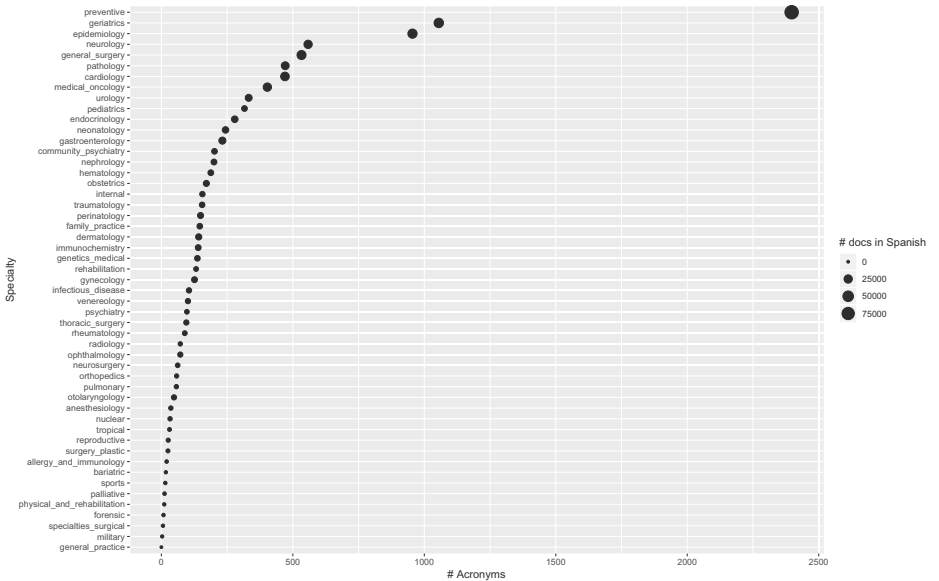
The resource⁴ generated is composed of 51 dictionaries with 3,603 acronyms identified from the Spanish subset of MEDLINE. Figure 2 illustrates the number of acronyms vs. the number of titles analyzed per specialty. The analysis of this relationship showed that out of the group of specialties with more than 1,000 titles available, pediatrics, rehabilitation, radiology, psychiatry, and orthopedics are the specialties with the higher ratio of acronyms per title. For instance, pediatrics has approximately one acronym in every 22 titles / abstracts.

We analyzed the ambiguity of acronyms by comparing the number of possible expansions in the corpus and their reduction when they were analyzed by specialty (Figure 3). The x axis shows the number of possible expansions in a specialty-independent inventory, while the y axis shows the average number of expansions in the specialties in which the acronym was used. The dot size represents the number of acronyms with the

³<https://biopython.org/>

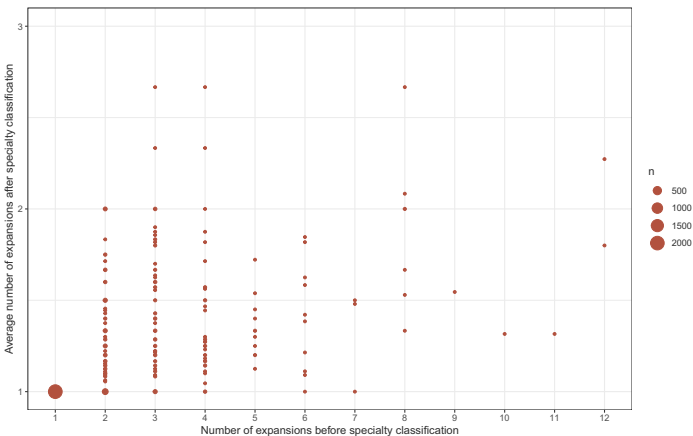
⁴Available at <https://github.com/plubeda/spanish-acronyms-pubmed>.

Figure 2. Acronyms per specialty and number of source document titles.

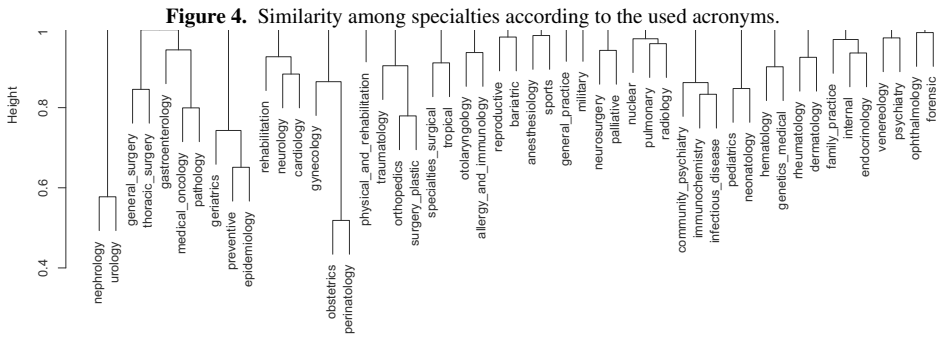


same behavior. We found that the acronyms DE (12), EC (12), CC (11), II (10) are the most ambiguous acronyms as they have ten or more possible expansions. We reduced the overall ambiguity from 1.32 to 1.06 senses/acronym, a reduction of 81% ambiguity (excess over 1) on average. Considering only ambiguous acronyms, we reduced it from 2.71 to 1.28 senses/acronym, a mean reduction of 84%.

Figure 3. Number of expansions before and after specialty classification.



Finally, Figure 4 depicts the similarity of the specialties with respect to the acronyms they use. The height at which any two specialties merge indicates the joining order: a smaller height represents a stronger relationship; e.g., “obstetrics” and “perinatology” are the most similar because they share 92% of their acronyms.



4. Discussion and Conclusion

The generated resource supported the hypothesis that creating a sense-inventory of acronyms classified by specialty reduces the level of ambiguity when compared to a specialty-independent sense inventory. Therefore, it can be used to improve contextual knowledge about an acronym and thus better disambiguate its meaning during the resolution process. The development of this resource also made it possible to identify the specialties in which more acronyms are used and those specialties with stronger links in terms of the acronyms used.

The proposed method can be expanded to any other language that has available resources in PubMed. In the future, we plan to use the generated resource for pre-processing electronic health records. We also would like to extract contextualized word embeddings for each acronym and make them available as features together with the dictionary resource.

References

- [1] S. Pakhomov, T. Pedersen and C.G. Chute, Abbreviation and acronym disambiguation in clinical discourse, *AMIA. Annual Symposium proceedings. AMIA Symposium* (2005), 589–593.
- [2] H. Yu, G. Hripcsak and C. Friedman, Mapping Abbreviations to Full Forms in Biomedical Articles, *Journal of the American Medical Informatics Association* **9**(3) (2002), 262–272. doi:10.1197/jamia.M0913.
- [3] H. Ao and T. Takagi, ALICE: an algorithm to extract abbreviations from MEDLINE, *Journal of the American Medical Informatics Association* **12**(5) (2005), 576–586.
- [4] H. Liu and C. Friedman, Mining terminological knowledge in large biomedical corpora, in: *Biocomputing 2003*, World Scientific, 2002, pp. 415–426.
- [5] W. Zhou, V.I. Torvik and N.R. Smalheiser, ADAM: another database of abbreviations in MEDLINE, *Bioinformatics* **22**(22) (2006), 2813–2818.
- [6] E. Adar, SaRAD: A simple and robust abbreviation dictionary, *Bioinformatics* **20**(4) (2004), 527–533.
- [7] A.S. Schwartz and M.A. Hearst, A simple algorithm for identifying abbreviation definitions in biomedical text, in: *Biocomputing 2003*, 2002, pp. 451–462.
- [8] A. Intxaurrenondo, M. Marimon, A. Gonzalez-Agirre, J.A. Lopez-Martin, H. Rodriguez Betanco, J. Santamaria, M. Villegas and M. Krallinger, Finding mentions of abbreviations and their definitions in Spanish Clinical Cases: the BARR2 shared task evaluation results, SEPLN, 2018.
- [9] F. Sánchez-León, ARBOREx: Abbreviation resolution based on regular expressions for BARR2, SEPLN, 2018.
- [10] S. Montalvo, R. Martínez, M. Almagro and S. Lorenzo, MANTRA-MED at Biomedical Abbreviation Recognition and Resolution-IberEval 2018., in: *IberEval@ SEPLN*, 2018, pp. 290–296.

Logical Rules and a Preliminary Prototype for Translating Mortality Coding Rules from ICD-10 to ICD-11

Vincenzo DELLA MEA^{a,1}, Mihai H POPESCU^a, Francesco GRIPPO^b, Chiara ORSI^b,
Friedrich HEUSER^c

^a*Dept. of Maths, Computer Science and Physics, University of Udine, Italy*

^b*ISTAT, Italy*

^c*Iris Institute, DIMDI, Germany*

Abstract. Iris is a system for coding multiple causes of death in ICD-10 and for the selection of the underlying cause of death, based on a knowledge base composed by a large number of rules. With the adoption of ICD-11, those rules need translation to ICD-11. A pre-project has been carried out to evaluate feasibility of transition to ICD-11, which included the analysis of the logical meta-rules needed for rule translation and development of a prototype support system for the expert that will translate the coding rules.

Keywords. Classifications, ICD-10, ICD-11, ontology mapping

1. Introduction

With its adoption by the 72th World Health Assembly in May 2019, ICD-11 (International Classification of Diseases, Revision 11) will become the new standard for coding diseases and health problems. For mortality statistics, the ICD coding is performed with the use of semi-automated coding systems, mainly Iris [1].

Iris is a system for coding multiple causes of death in ICD-10 and for the selection of the underlying cause of death. Iris is based on the international death certificate form and coding rules provided by WHO (World Health Organisation). The core component of Iris are the decision tables, currently based on ICD-10 codes. For the transition of Iris to ICD-11, it will be necessary to translate the decision tables and to allow them to include all the features of the new revision.

This transition is expected to have a big impact on the system transition and use. Periodic revision of the ICD is essential to stay abreast of advances in medical science and changes in medical terminology [2,3]. Institutionally, revision of the ICD requires an enormous investment of national resources to revise software, training, publications, edit procedures, etc [4]. For the Iris transition to ICD-11, classification and coding experts are needed. To support their work, formal procedures are needed to ensure the correctness of the transition and validation of the system.

¹ Corresponding Author, Vincenzo Della Mea, Dept. Of Mathematics, Computer Science and Physics, University of Udine, via delle Scienze 206 – 33100 Udine Italy E-mail: vincenzo.dellamea@uniud.it.

A pre-project has been carried out to evaluate feasibility of transition to ICD-11, which included the analysis of the logical meta-rules needed for rule translation and development of a prototype translation support system aimed at the expert [5]. This paper describes the logical rules identified for translating from ICD-10 to ICD-11 with their consequences from the point of view of automation and human expert intervention, with an overview of the prototype, and evaluation of efforts needed.

1.1. Decision tables

The decision tables form a knowledge base of relations between pairs of codes (representing the causes of death reported on the death certificate) that must be taken into consideration during the application of the steps for the selection of the underlying cause.

This knowledge base was first developed by the NCHS (US National Center for Health Statistics) for the ACME system [6]. Successively it has been embedded in the new automated coding system Iris and updated on the basis of the recommendations of the Mortality Reference Group, which operates in the network of the WHO Collaborating centers for the Family of International Classifications (WHO-FIC) [7].

In practical terms, the decision tables are a list of possible kind of relations between pairs of terminal code. The relation between the codes are specified as “rule type”, which are used in the different steps of selection.

1.2. Mapping between classification versions

Along with ICD-11, the WHO releases the mapping table between the ICD-10 and ICD-11 classification. The mapping table contains the mapping of both terminal codes and higher categories which can be used to add detail.

The mapping between the ICD-10 and ICD-11 entities can be specified by cardinality as: 1×1 (equivalent), $1 \times n$, $n \times 1$ and $n \times n$ (structure of the classification change and the categories intersect each other) with axb where a is the cardinality of the ICD-10 codes and b is the cardinality of ICD-11 codes.

2. Methods

For this analysis the 2019 decision tables are considered, since they include the most recent ICD-10 updates. The most represented rule is the DUETO, with more than 27 million records (90.1% of the total rules). The analysis thus focused on the DUETO rules. These rules represent the vast majority of the rules and are the most critical for coding.

2.1. Translation method

From former evaluations of the transition between the ICD-9 and ICD-10, it is known that the transition to ICD-11 will have an impact and the mapping will not be enough to completely automate the transition. The mapping table gives the possibility to translate single codes, but when it comes to the rules, we need to translate relationships between codes. Due to the relation between the pair elements, we cannot translate codes separately, but we need to interpret the relation from a logical point of view and the impact that the diverse mapping cardinality has on the two codes involved.

2.2. DUETO rule-type translation

The *DUETO* relation can shortly be described as:

Code A DUETO B if B is an acceptable cause of A (according to ICD provisions); in IRIS terminology, A is called codeDef and B is called subcodeDef.

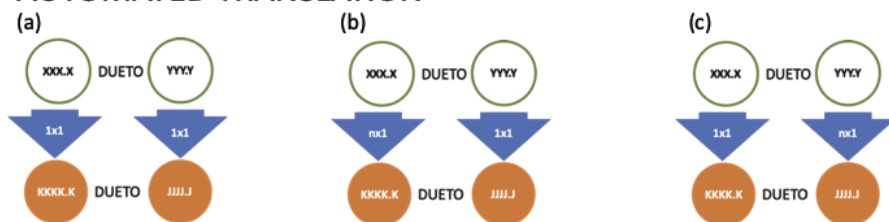
The goal of the translation is to identify all the possible translation rules that can be automated. Afterwards we define the *DUETO* rule-type in combination with the mapping cardinality of the two codes that can be automated, and which needs expertise support.

The decision tables are based on pair of codes, but since the *DUETO* rules may include many consecutive pairs, the conditions can be represented also as ranges of codes. Using ranges, it is possible to achieve better results but also to reuse part of the translation of the *subcodeDef*.

Given a rule, the basic idea is to verify whether a mapping exists between single codes for both *codeDef* and *subcodeDef* and of which kind. A rule can be translated automatically if both sides can be translated automatically. In some cases, we may need a different translation of codes for *codeDef* and *subcodeDef*, since the consequences of the rule-type could be different. So we want to always suggest the translation codes based on the mapping but, basing on the mapping type and the implications of the relation, the system also suggest which are the codes that can be accepted or need expert supervision.

Since the disease knowledge evolves, some ICD-10 codes may have no mappings to the ICD-11 classification since the concept is no longer used, or ICD-11 codes may have no mapping from the ICD-10 since they are new. Those codes are problematic and need expert supervision. Sometimes, using ranges for the translation, some of those codes could be handled. Since the classification is hierarchical, translating the higher levels could avoid manual intervention on some of the terminal codes with no mapping, since the results would be covered by the parent translation.

AUTOMATED TRANSLATION



TRANSLATION WITH EXPERT INTERVENTION



Figure 1. Summary of translation cases.

3. Results

Basing on the cardinality of the mappings on both sides, we can distinguish between manual and automatic translations.

The equivalent mapping ($1x1$), when found on both sides of the rule, is the basic case of automation, applicable to all the rule types. On the other side, when a mapping with cardinality of nxn is present, translation will always be manual. Rules where mapping of type $nx1$ and $1xn$ appear are those that provide some chance of automated translation, as depicted in Figure 1.

The key for the DUETO rule translation is to remember that is not representing causality, but the possibility of causality. Thus, whenever aggregated conditions are present in the ICD-11 translation, such causality could not be excluded and thus the mapped rule is valid as obtained from the mappings. Let’s examine two examples of automated translation in table 1.

Table 1. Automated translation examples.

case	ICD-10 rule	Mappings involved	ICD-11 rule
(a)	A01.0 (Typhoid fever) DUETO C33 (Malignant neoplasm of trachea)	A01.0 \equiv 1A07 C33 \equiv 2C24	1A07 (Typhoid fever) DUETO 2C24 (Malignant neoplasms of trachea)
(b)	F03 (Unspecified dementia) DUETO R54 (Senility) <i>But also:</i> F01 (Vascular dementia) DUETO R54 (Senility)	F03 \sqsubseteq 6D8Z R54 \equiv MG2A <i>But also:</i> F00 (Dementia in Alzheimer disease) \sqsubseteq 6D8Z F01 (Vascular dementia) \sqsubseteq 6D8Z	6D8Z (Dementia, unknown or unspecified cause) DUETO MG2A (Old age)
(c)	J96.9 (Respiratory failure, unspecified) DUETO F03	J96.9 \equiv CB41.2 F03 \sqsubseteq 6D8Z <i>But also:</i> F00 \sqsubseteq 6D8Z F01 \sqsubseteq 6D8Z	CB41.2 (Respiratory failure, unspecified as acute or chronic) DUETO 6D8Z

Let’s now examine the cases that need expert intervention in table 2. In this case, since ICD-11 conditions are recorded with greater detail, we cannot ensure all of them are reasonably part of a DUETO rule (but the expert will know and decide). No example is provided for mixed situations involving nxn because too complex to be shown here.

Table 2. Examples of rules translation needing expert intervention.

case	ICD-10 rule	Mappings	ICD-11 rule
(d)	I46.9 (Cardiac arrest, unspecified) DUETO R26.3 (Immobility)	I46.9 \supseteq MC82 (Cardiac arrest) I46.9 \supseteq MC82.0 (Ventricular tachycardia and fibrillation cardiac arrest) I46.9 \supseteq MC82.1 (Bradycardic cardiac arrest) I46.9 \supseteq MC82.2 (Asystolic cardiac arrest) I46.9 \supseteq MC82.3 (Cardiac arrest with pulseless electrical activity) R26.3 (Immobility) \equiv MB44.3 (Immobility)	We cannot ensure that all types of cardiac arrest can be due to immobility
(e)	I27.9 (Pulmonary heart disease, unspecified) DUETO B44.1 (Other pulmonary aspergillosis)	I27.9 \equiv BB0Z (Pulmonary heart disease or diseases of pulmonary circulation, unspecified) B44.1 \supseteq CA82.4 (Aspergillus-induced allergic or hypersensitivity conditions) B44.1 \supseteq 1F20.12 Chronic pulmonary aspergillosis)	We cannot ensure that both CA82.4 and 1F20.12 can cause Pulmonary heart disease.

In cases like (d) and (e), the expert should be given the list of mappings that map to the same ICD-11 entity, to let him decide which ones generate valid DUETO rules.

Using this method for the translation the expert has a tool to quickly get an overview of the suggestions for the translation, but also has the possibility to focus more on the results that are suggested as manual after the validation of the method.

CodeDef and *subcodeDef* can be translated separately, and for the translation of the *subcodeDef* we can reuse the same translation in different rules. A further improvement is the translation reuse of a sub-range, that can be used for the translation of ranges that incorporate the translated range. Both techniques help in reducing human intervention.

Further details on the translation methodology can be found in [5].

By exploring the distribution of mapping types in the coding rules, we evaluated in about 3.2 million pairs those needing manual intervention, out of about 27 millions. With the reuse of the validated range translation we estimated a relevant reduction, where the cases that still need manual support are about 87000 codes (considering manual translation of *codeDef* and *subcodeDef*).

3.1. Prototype

The prototype is aimed at providing decision support in the translation of mortality rules from ICD-10 to ICD-11. For the implementation, we choose a web-based model where experts can work collaboratively from a different location but also for the consistency of the results. It has a full implementation of the visualization and editing of the “DUETO” translation rules of the decision tables. Rules are grouped where possible with the same *subcodeDef* to ease the translation and facilitate maintenance over time.

4. Conclusion

The presented method and prototype seem suitable for supporting the process of transition of Iris from ICD-10 to ICD-11, however it further needs expert validation to correctly estimate the workload needed. Adaption of results obtained on the DUETO rules is being carried out for the other rules, in particular those describing direct sequels of conditions.

References

- [1] The Iris Institute. URL: www.iris-institute.org
- [2] T.Boerma, J.Harrison, R.Jakob, C.Mathers, A.Schmider, S.Weber, Revising the ICD: explaining the WHO approach, *Lancet* **388** (2016), 2476-2477.
- [3] The World Health Organization. *ICD-11 Implementation or Transition Guide*. WHO, 2019. URL: https://icd.who.int/docs/ICD-11%20Implementation%20or%20Transition%20Guide_v105.pdf
- [4] R.N. Anderson, A.M. Miniño, D.L. Hoyert, H.M. Rosenberg, Comparability of Cause of Death Between ICD-9 and ICD-10: Preliminary Estimates, *National Vital Statistics Report* **49** (2001),1-32.
- [5] The Iris Institute. *Evaluation Project on Iris and ICD-11*. URL: <https://www.dimdi.de/dynamic/en/classifications/iris-institute/iris-and-icd-11/>
- [6] R.A. Israel, Automation of mortality data coding and processing in the United States of America, *World Health Stat Q* **43** (1990),259-62
- [7] S.Navarra, M.Cappella, L.A.Johansson, L.Pelikan, L.Frova, F.Grippo, *Decision Table Editor: a web application for the management of the international tables for mortality coding*, ISTAT working papers, Rome, 6/2016. URL: <https://www.istat.it/it/archivio/184113>

Machine Learning Assisted Citation Screening for Systematic Reviews

Anjani DHRANGADHARIYA^{a1}, Roger HILFIKER^b, Roger SCHAER^a and Henning MÜLLER^{a,c}

^a *University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland*

^b *School of Health Sciences, HES-SO Valais-Wallis, Leukerbad, Switzerland*

^c *University of Geneva (UNIGE), Geneva, Switzerland*

Abstract. Evidence-based practice is highly dependent upon up-to-date systematic reviews (SR) for decision making. However, conducting and updating systematic reviews, especially the citation screening for identification of relevant studies, requires much human work and is therefore expensive. Automating citation screening using machine learning (ML) based approaches can reduce cost and labor. Machine learning has been applied to automate citation screening but not for the SRs with very narrow research questions. This paper reports the results and observations for an ongoing research that aims to automate citation screening for SRs with narrow research questions using machine learning. The research also sheds light on the problem of class imbalance and class overlap on the performance of ML classifiers when applied to SRs with narrow research questions.

Keywords. Systematic reviews, Automation, Natural language processing, Machine learning

1. Introduction

Summarizing evidence in a specific domain through the process of conducting systematic reviews (SR) becomes increasingly difficult with an exponential increase in the number of primary research publications and an increasing variety of publishers. With this increased publication rate, citation screening for conducting the SRs has become time-consuming and labour intensive. The core of citation screening involves manually categorizing the studies found in literature search into relevant or non-relevant depending upon predefined criteria for the research question at hand [1]. This screening is usually performed by two independent reviewers, who often cannot keep up with the manual process of screening the studies and constantly updating outdated SRs [2].

In the area of physiotherapy and rehabilitation such an exponential increase in the number of publications is also observed². For an ongoing review on exercise and non-exercise interventions in reducing cancer-related fatigue retrieved over 30,000 references, about 2,000 of which were published in 2017 alone. It took more than 200 hours for each of the two independent reviewers to manually assess the titles and abstracts

¹ Corresponding Author: Anjani Dhrangadhariya, University of Applied Sciences Western Switzerland (HES-SO), Technopole 3, 3960 Sierre, Switzerland; E-mail: anjani.dhrangadhariya@hevs.ch.

² [https://www.ncbi.nlm.nih.gov/pubmed?term=\(physiotherapy\)%20OR%20rehabilitation](https://www.ncbi.nlm.nih.gov/pubmed?term=(physiotherapy)%20OR%20rehabilitation)

for relevance to the research question before the studies were taken for further meta-analysis [3]. Supervised machine learning based classification approaches are successfully applied for automation of citation screening but either only for broad and shallow SRs [4] or SRs that retrieved fewer than 6,000 studies [5]. However, there are SRs that address very specific research questions leading to narrow, predefined criteria for selection of relevant studies to be included for meta-analysis. For such narrow SRs, inclusion prevalence becomes as low as 10%, which means that out of all the studies retrieved during the search phase, 90% are excluded as non-relevant. A narrow research question combined with a low inclusion prevalence leads to class imbalance and class overlap problems for classification tasks that generally reduce classifier performance [6]. Class overlap cannot be artificially controlled but class imbalance can be tackled using oversampling or undersampling [7]. In this work, we aim to explore machine learning and natural language processing to assist citation screening in SRs with a narrow research question and low inclusion prevalence using word embeddings and random oversampling.

2. Methods

2.1 Datasets

Data sets from the open access literature were used in the work is described below.

- a) *Word embedding* Titles and abstracts (TA) of 2.09 million studies included in the PubMed Central Open-Access subset (PMC-OA) were used to generate semantic word embedding using the two most common architectures: word2vec and fastText [8,9].
- b) *Data for the screening process* The data set used to test automation approaches included the studies identified for citation screening in an update to the systematic review by Hilfiker *et al.* [3]. The data set included TA from 31,279 studies identified during the search phase of this SR. These studies were already manually assessed for relevance and labelled by two reviewers into two mutually exclusive labels. 4,066 studies assessed as relevant were labelled “include” and the rest were labelled “exclude”. The inclusion prevalence for this case is only about 13% leading to class imbalance.

2.2 Screening Automation

To automate the screening and explore the effect of class imbalance, six supervised binary classifiers were trained to classify documents into “include” (relevant) vs. “exclude” (non-relevant) using the data set of Section 2.1 represented using corpus-specific static word embedding. This approach follows steps enumerated below.

1. *Word embedding generation* To generate embedding, the PMC-OA data were lower-cased and all punctuation except the hyphens were removed. Phrase generation was then performed using the word2phrase tool³ to identify frequently occurring bi-grams. The output of phrase generation along with the unigrams was fed to gensim’s word2vec⁴ and to fastText⁵ using the

³<https://github.com/travisbrady/word2phrase>

⁴<https://radimrehurek.com/gensim/models/word2vec.html>

hyperparameters in Table 1 to obtain two dense, semantic, real-valued word embeddings [10].

Parameter	Value	Parameter	value	Parameter	value
Size	300	Alpha	0.05	bucket*	2000000
Window	5	min alpha	0.0001	minn*	3
min eount	5	sample	0.0001	maxn*	6
sg	1	iter/ epoch	5	seed	1
hs	1	negative	5		

Table 1. Hyperparameter values used to generate word embeddings using gensim’s word2vec and the fastText functionality. (*) means that these parameters were available only for the fastText embeddings.

2. *Data set for text pre-processing* Post deduplication and removal of non-English language studies led to 25,540 studies remaining. For these remaining studies, the text was lower-cased and tokenized into words using NLTK (Natural Language ToolKit). Irrelevant tokens were removed using a predefined set of stop-words provided by NLTK and PubMed. Additional corpus-specific stop-words identified during the experiments with the training set were removed accordingly. The text normalization process converted British English terms into American English. After token lemmatization, a corpus vocabulary was constructed from all the unique unigram and bigram tokens. To scale this vocabulary down, uninformative short tokens with fewer than five characters and the tokens with vocabulary count lower than five were removed assuming they were not representative of the classes.
3. *Random oversampling* Class imbalance often deteriorates the classifier performance, so in the present data set it was addressed using naive random oversampling. This method randomly duplicates data points from the minority class and brings the total number of instances in the minority class equal to the majority class [6].
4. *Feature extraction* Feature extraction was performed to generate real-valued, dense feature vectors from text. These features were used as input to train the supervised ML classifiers. For non-neural ML classifiers, from each of the 25,540 studies word vectors were extracted for each token using both generated word embeddings. All vectors for each token in the individual study were averaged over the entire study normalized by study length to produce a single 300-dimensional vector representing a particular study. For the Convolutional Neural Network (CNN), feature extraction was part of the model, where a static, non-trainable weight matrix generated from the word embedding was provided with the embedding layer to extract token word vectors during training.
5. *Classifier training and evaluation* Six classifiers including Logistic Regression (LR), Support Vector Machines (SVM), k-nearest neighbor (KNN), Decision Trees-CART (DT), Random Forest (RF), and CNNs were trained and evaluated over the generated text feature to perform binary classification. Hyperparameter tuning was performed using GridSearch. A CNN was trained using hyperparameters loosely based on suggestions from Zhang *et al.* [11]. Classifier performance before and after oversampling was evaluated on an unseen validation data set and metrics pertinent to imbalanced classification

⁵ <https://radimrehurek.com/gensim/models/fasttext.html>

like precision, recall, F1 and precision-recall AUC (PR-AUC) score were tracked[7].

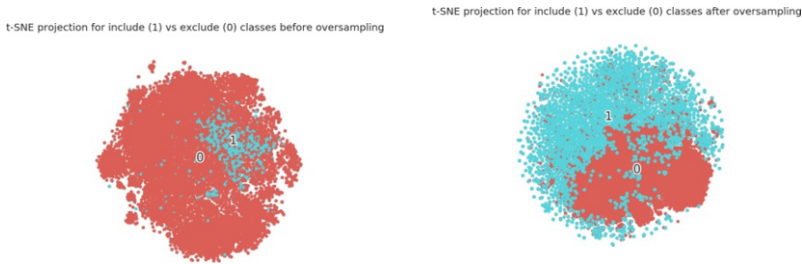


Figure 1. A t-SNE projection for 25,540 studies labelled “include” vs. “exclude” before and after oversampling.

3. Results and Discussion

The data set used for training and evaluating the classifiers comprised 2,259 relevant studies labelled “include” and 23,281 labelled “exclude”. Cosine similarity between class centroid vectors for the “include” and “exclude” classes before and after oversampling is 0.985814 and 0.985824 respectively. A high cosine similarity is indicative of high class overlap even after tackling the class imbalance using oversampling (see Figure 1).

The results obtained by applying classifiers on the data set before and after random oversampling are summarized in Table 2 and 3. Before oversampling, the classifiers focused on improving the performance for the majority class but in reality they are simply predicting the majority class as noticeable from the relatively high F1 score for the exclude class. Upon oversampling, the overall classifier performance drastically improves for the minority class especially the precision (see Table 3), while the precision for the majority class is reduced with a small improvement in recall.

Times ne	embedding	Class “include”				Class “exclude”		
		Precision	Recall	F1	PR-AUC	Precision	Recall	F1
LR	fastText	0.4044	0.8990	0.5576	0.6008	0.9891	0.8746	0.9283
SVM	fastText	0.6538	0.4640	0.5428	0.6317	0.9463	0.9746	0.9602
KNN	word2vec	0.6536	0.6066	0.6288	0.6512	0.9619	0.9685	0.9652
DT	fastText	0.2961	0.8627	0.4394	0.4287	0.9837	0.8000	0.8821
RF	fastText	0.4995	0.7892	0.6108	0.5921	0.9780	0.9209	0.9485
CNN	fastText	0.6545	0.5032	0.5690	0.6388	0.9511	0.9732	0.9620

Table 2. Classifier performance before random oversampling for the “include” and “exclude” classes

If the task is considered a classification task, a high class overlap still leads to unacceptable precision and recall values for citation screening implying the inclusion of false-positive studies and exclusion of false-negative studies. Our future work has two directions: Firstly, experimenting with systematic oversampling techniques like Synthetic Minority Over-Sampling Technique (SMOTE) and Adaptive Synthetic (ADASYN) and considering this task as data extraction rather than classification [12].

Classifier	embedding	Class “include”				Class “exclude”		
		Precision	Recall	F1	PR-AUC	Precision	Recall	F1
LR	word2vec	0.8981	0.8850	0.9116	0.9342	0.8951	0.9090	0.8816
SVM	fastText	0.8914	0.8818	0.9012	0.9378	0.8990	0.8792	0.8890
KNN	word2vec	0.8860	0.8303	0.9500	0.9321	0.9418	0.8055	0.8682
DT	fastText	0.8348	0.8201	0.8510	0.8734	0.8285	0.8463	0.8126
RF	word2vec	0.8695	0.8918	0.8488	0.9279	0.8966	0.8757	0.8560
CNN	word2vec	0.9034	0.7480	0.8183	0.9318	0.7850	0.9200	0.8471

Table 3. Classifier performance after random oversampling for the “include” and “exclude” classes.

4. Conclusion

To the best of our knowledge, this is the first attempt to explore citation screening automation for a narrow systematic review topic using domain-specific word embedding on a range of ML classifiers. The research specifically sheds light on the impact of class imbalance and class overlap on the classifier performance before and after oversampling as also discussed by [6,7]. Knowledge of these challenges is useful for further development of automation approaches for citation screening.

References

[1] J. Higgins and S. Green, *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration, Mar 2011.

[2] S. Khangura, K. Konnyu, R. Cushman, J. Grimshaw, and D. Moher, “Evidence summaries: the evolution of a rapid review approach,” *Syst Rev* **1**, p. 10, Feb 2012.

[3] R. Hilfiker, A. Meichtry, M. Eicher, L. Nilsson Balfe, R. H. Knols, M. L. Verra, and J. Taeymans, “Exercise and other non-pharmaceutical interventions for cancer-related fatigue in patients during or after cancer treatment: a systematic review incorporating an indirect-comparisons meta-analysis,” *British Journal of Sports Medicine* **52**(10), pp. 651–658, 2018.

[4] A. Bannach-Brown, P. Przyby?a, J. Thomas, A. S. C. Rice, S. Ananiadou, J. Liao, and M. R. Macleod, “Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error,” *Syst Rev* **8**, p. 23, Jan 2019.

[5] I. Lerner, P. Cr?equit, P. Ravaud, and I. Atal, “Automatic screening using word embeddings achieved high sensitivity and workload reduction for updating living network meta-analyses,” *J. Clin. Epidemiol.* **108**, pp. 86–94, Apr. 2019.

[6] V. Garc?a, R. Alejo, J. S. S?nchez, J. M. Sotoca, and R. A. Mollineda, “Combined effects of class imbalance and class overlap on Instance-Based classification,” in *Intelligent Data Engineering and Automated Learning – IDEAL 2006*, pp. 371–378, Springer Berlin Heidelberg, 2006.

[7] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLoS ONE* **10**(3), p. e0118432, 2015.

[8] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality.,” *NIPS’13 Proceedings of the 26th International Conference on Neural Information Processing Systems* **2**, pp. 3111–3119, Dec 2013.

[9] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *Transactions of the Association for Computational Linguistics* **5**, pp. 135–146, 2017.

[11] Y. Goldberg, “A primer on neural network models for natural language processing,” Oct. 2015.

[12] Y. Zhang and B. Wallace, “A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification,” Oct. 2015.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and others, “SMOTE: synthetic minority over-sampling technique,” *J. Artif. Organs* , 2002.

Machine Learning Explainability in Breast Cancer Survival

Tom JANSEN^{a,b}, Gijs GELEIJNSE^b, Marissa VAN MAAREN^{b,c},
Mathijs P. HENDRIKS^{b,d}, Annette TEN TEIJE^a, and Arturo MONCADA-TORRES^{b,1}

^aDept. of Computer Science, Vrije Universiteit Amsterdam, NL

^bNetherlands Comprehensive Cancer Organization (IKNL), Eindhoven, NL

^cUniversity of Twente, Enschede, NL

^dDept. of Medical Oncology, Northwest Clinics Alkmaar, NL

Abstract. Machine Learning (ML) can improve the diagnosis, treatment decisions, and understanding of cancer. However, the low explainability of how “black box” ML methods produce their output hinders their clinical adoption. In this paper, we used data from the Netherlands Cancer Registry to generate a ML-based model to predict 10-year overall survival of breast cancer patients. Then, we used Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) to interpret the model’s predictions. We found that, overall, LIME and SHAP tend to be consistent when explaining the contribution of different features. Nevertheless, the feature ranges where they have a mismatch can also be of interest, since they can help us identifying “turning points” where features go from favoring *survived* to favoring *deceased* (or vice versa). Explainability techniques can pave the way for better acceptance of ML techniques. However, their evaluation and translation to real-life scenarios need to be researched further.

Keywords. Artificial Intelligence, interpretability, oncology, prediction model

1. Introduction

Although it has been shown that Machine Learning (ML) methods are able to predict oncological outcomes [1], there are still a few factors that hinder their widespread clinical adoption. One of these factors is the lack of trust in the models. Often, ML tools are considered black boxes. If decisions need to be made that are based (at least partially) on predictions made by ML algorithms, users need to be able to understand how and why the algorithm has come up with that decision [2].

In the last couple of years, ML explainability has gained considerable interest. Recently, two techniques have been proposed and developed to make ML models more interpretable: Local Interpretable Model-Agnostic Explanations (LIME) [3] and SHapley Additive exPlanations (SHAP) [4]. However, evaluation of these tools remains relatively unexplored. Although some studies have attempted to evaluate explanations of

¹Corresponding author: Arturo Moncada-Torres; Zernikestraat 29, 5612 HZ Eindhoven, NL;
E-mail: a.moncadatorres@iknl.nl.

model predictions by letting users test which explanations are more understandable or intuitive [5], more analytical, standardized evaluations are needed.

In this paper, we used data from the Netherlands Cancer Registry (NCR) to generate a predictive model of 10-year overall survival (OS) after curative breast cancer surgery. Then, we applied LIME and SHAP to the obtained model to explain its predictions. Finally, we evaluated said interpretability methods by analysing their explanations and the agreement between them, which allowed us to identify features' turning points.

2. Materials & Methods

We used NCR data granted under data request K18.999. It consisted of demographic, clinical, and pathological data of patients in the Netherlands diagnosed between 2005 and 2008 with non-metastatic breast cancer who underwent surgery. Features included age, tumor characteristics, hormonal receptor statuses, clinical and pathological TNM-staging, and number of removed and positive lymph nodes. We imputed missing values using Deep Learning and K-Nearest Neighbor (KNN). We defined 10-year OS as the target variable for our model. The final dataset consisted of 46,284 patients and 31 features.

We performed feature selection using a combination of 21 different filter and wrapper methods. Each of them output a ranking ordering feature predictiveness. Then, we computed the median of these rankings and chose the six best ranked features: 1. *age*, 2. ratio between the number of positive and removed lymph nodes (*ratly*), 3. number of removed lymph nodes (*rly*), 4. tumor size in millimeters (*ptmm*), 5. pathological TNM stage (*pts*), and 6. tumor grade (*grd*).

We experimented with a variety of ML tools: Random Forest, Extreme Gradient Boosting (XGB), KNN, Artificial Neural Networks, Naïve Bayes, and Logistic Regression. We performed a randomized grid search to train, test, and optimize each of them. Since our target variable had a class distribution of roughly 75% (*survived*) versus 25% (*deceased*), we used stratified 10-fold cross-validation to optimize the models' hyperparameters. Then, we evaluated their performance using the Area under the Curve (AUC) as a metric, with XGB yielding the highest value (0.78). Therefore, we used XGB for the rest of this study.

In order to better understand the model's predictions, we used LIME and SHAP. On the one hand, LIME approximates *individual predictions* of a (black box) model with a *local* (interpretable) surrogate model that is as close as possible to the original one. Explanations are produced by minimizing a loss function between the predictions of both of them. The complexity of the surrogate model is used to explain the original model [3]. On the other hand, besides offering local interpretability, SHAP allows to explain a model *globally* by expressing it as linear functions of features [4]. In other words, it explains how much the presence of a feature contributes to the model's *overall predictions*.

To assess the consistency of LIME in explaining individual predictions, we applied it to each of the predictions of the test set (20% of the data) 100 times. We defined consistency when LIME assigned values with the same sign in all cases. Then, we evaluated global variable importance using SHAP. Finally, we tested agreement between LIME and SHAP values by comparing their instances (i.e., local) explanations in the test set. We defined agreement when both methods assigned either a positive or a negative contribution to the same feature of the same data instance.

3. Results

Figure 1 shows five representative data instances (i.e., patients) picked at random of the LIME consistency test. The x -axis shows the values for a particular feature, while the y -axis denotes the feature weight assigned to that value by LIME. A positive weight means it contributed to *survived*, while a negative one contributed to *deceased*. Across all plots, the position of each box is the same for each patient. For instance, the first box of each plot corresponds to the same patient, who was 51 years old, had a *ratly* value of 0, a *rly* value of 1, etc. LIME was consistent in >98% of the cases of *age*, *pts*, and *grd*. However, for *ratly*, *rly*, and *ptmm* LIME was consistent in 74%, 8%, and 55% of the cases, respectively. These cases correspond to the boxes in Figure 1 that cross 0 (dotted line), which means that LIME yielded contradictory weights.

Figure 2 shows the SHAP values of the six features used by the global model. The x -axis shows the SHAP values that correspond to the features shown on the y -axis. The color scale indicates the feature values, which range from low (blue) to high (red). Similarly to LIME, a positive SHAP value contributes to *survived*, while a negative one contributes to *deceased*.

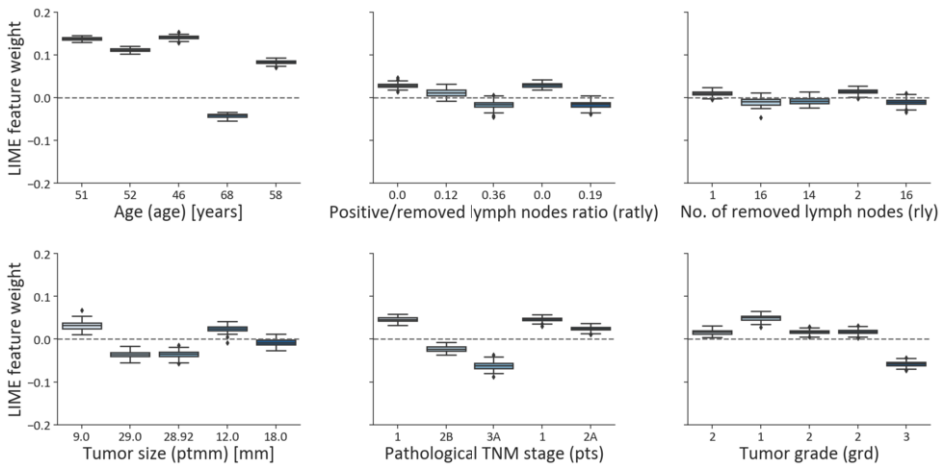


Figure 1. LIME consistency of five representative data instances (i.e., patients) picked at random. Contradictory explanations correspond to boxes that cross the dotted line at 0. Practically, *age*, *pts*, and *grd* had consistent LIME values, while the opposite can be said for *ratly*, *rly*, and *ptmm*.

The percentage of instances where LIME and SHAP agreed on their explanations for each feature was as follows: *age*, 97.5%; *ratly*, 95.9%; *rly*, 87.8%; *ptmm*, 91.9%; *pts*, 99.6%; *grd*, 99.9%. Figure 3 shows the individual feature weights for *age* (since it was the best-ranked feature) and for *rly* (since it was the feature with the largest disagreement). The x -axis denotes the feature values, while the y -axis denotes the feature weights assigned by the interpretability methods. Circles indicate an agreement between them, while crosses indicate the opposite.

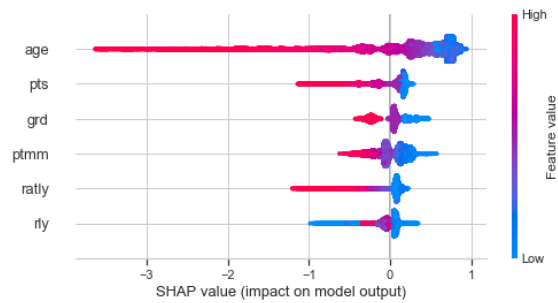


Figure 2. Summary plot of all SHAP values. Globally, *age* has the biggest impact on the model output.

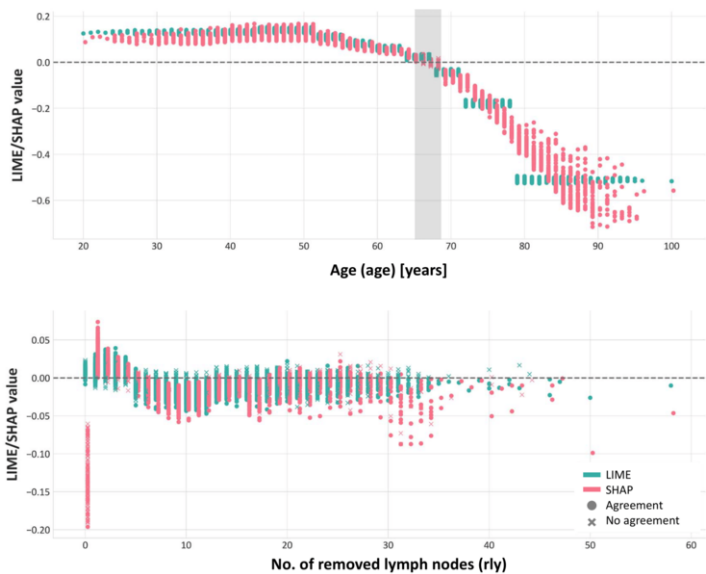


Figure 3. Feature weights assigned by LIME and SHAP to all test set instances for *age* and *rly*. The shaded area shows clear *age*’s clear “turning point”. SHAP markers were shifted slightly along the x-axis for clarity.

4. Discussion

Figure 1 shows that LIME tends to assign consistent values to categorical features. For example, a *pts* of 1 is assigned almost identical feature weights in different patients. A similar thing occurs for a *grad* of 2. However, LIME has more difficulties with numerical features. For instance, there is very little difference in the impact that having 1 or 16 lymph nodes removed has on the model predictions. We think this is because LIME discretizes continuous features by binning them and treating them as categorical, losing information. Figure 1 also shows that LIME weights can be contradictory. For example, in the *rly* case, LIME values for the same patient are often inconsistent. It has been suggested that LIME’s uncertainty can be explained by randomness in the sampling procedure and the variation of interpretation quality across different data instances [6], which is in line with the presented results.

Figure 2 combines the features' effects (x-axis) with their importance (y-axis). At a global level, *age* is the most important feature, while *rly* is the least important. This could be explained by its non-monotonic behaviour (i.e., low *rly* values are assigned positive and negative weights).

Although LIME and SHAP values show a similar trend overall for both *age* and *rly* (Figure 3), we can also distinguish specific regions of mismatch. These are of particular interest, since they can help us identify "turning points" in the features' values. For example, in the case of *age*, mismatches occur approximately between 65 and 68 years (shaded area). This could explain where the model considers *age* to contribute towards *survived* or towards *deceased*.

5. Conclusion

In this study, we used breast cancer data from the NCR to generate an XGB-based model for predicting 10-year OS. We explained the model's predictions using LIME and SHAP and compared their performance. In few cases, LIME showed inconsistent and contradictory explanations of individual predictions. Furthermore, comparing LIME and SHAP showed agreement between them in 95.4% of the instances. The regions of mismatch allowed us to identify "turning points" in the features' values, which indicate where features go from favoring *survived* to favoring *deceased* (or vice versa).

Methods like LIME and SHAP are a first step to provide a more interpretable way of explaining complex models than what the models are capable of themselves. It is important to keep in mind that perfect explanations are also infeasible, since there is no gold standard to which the explanations can be compared. This also makes the evaluation of these methods a challenge. These type of methods pave the way for larger use and acceptance of ML techniques for digital health applications. However, their evaluation and translation to different fields need to be researched further.

References

- [1] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal*, vol. 13, pp. 8–17, 2015.
- [2] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable AI systems for the medical domain?" *Preprint arXiv:1712.09923*, 2017.
- [3] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, ACM, 2016, pp. 1135–1144.
- [4] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. in Neural Information Processing Systems*, 2017, pp. 4765–4774.
- [5] A. A. Freitas, "Comprehensible classification models: A position paper," *ACM SIGKDD explorations newsletter*, vol. 15, no. 1, pp. 1–10, 2014.
- [6] H. Fen, K. Song, M. Udell, Y. Sun, Y. Zhang, et al., "Why should you trust my interpretation? Understanding uncertainty in LIME predictions," *Preprint arXiv:1904.12991*, 2019.

Machine Learning for Automatic Encoding of French Electronic Medical Records: Is More Data Better ?

Julien GOBEILL^{a,b,1}, Patrick RUCH^{a,b}, and Rodolphe MEYER^c

^a*SIB Text Mining group, Swiss Institute of Bioinformatics, Geneva, Switzerland*

^b*HES-SO / HEG, Information Sciences, Geneva, Switzerland*

^c*Information Systems Department, University Hospitals of Geneva (HUG), Geneva, Switzerland*

Abstract. The encoding of Electronic Medical Records is a complex and time-consuming task. We report on a machine learning model for proposing diagnoses and procedures codes, from a large realistic dataset of 245 000 electronic medical records at the University Hospitals of Geneva. Our study particularly focuses on the impact of training data quantity on the model's performances. We show that the performances of the models do not increase while encoded instances from previous years are exploited for learning data. Furthermore, supervised models are shown to be highly perishable: we show a potential drop in performances of around -10% per year. Consequently, great and constant care must be exercised for designing and updating the content of such knowledge bases exploited by machine learning.

Keywords. Medical coding, machine learning, text mining.

1. Introduction

The coding, or encoding, of medical records with standard classifications serves multiple purposes for healthcare actors, including billing and reimbursement, quality control, epidemiological surveillance, or cohort identification for clinical trials [1,2]. This task is usually performed by professional coders, at the term of an episode of care (EOC). These coders have to master knowledge in the field of medicine, along with the handling of large code sets, coding rules and local guidelines [3]. For these reasons, manual coding is expensive and time-consuming. The amount of available electronic medical records (EMRs), potentially exploitable by computers, has steadily grown in recent years [4]. Hence, machine learning algorithms are today likely to produce accurate and effective tools for assisting human coders [5].

We report on the development and evaluation of a machine learning – or supervised – model for suggesting codes, from a large dataset of EMRs.

A citation attributed to Peter Norvig, Google's Chief Scientist, claims that "more data beats clever algorithms" [6]. Several reported works on text categorization tend to support this idea, and incite to direct efforts towards increasing the size of annotated

¹ Corresponding Author, SIB Swiss Institute of Bioinformatics, quartier Sorge - bâtiment Amphipôle, 1015 Lausanne, Switzerland; E-mail: julien.gobeill@sib.swiss.

learning collections [7,8]. Our study has particularly focused on the impact of exploiting growing amounts of learning data on the algorithm's performances. The figure below illustrates the global workflow of the automatic encoding process.

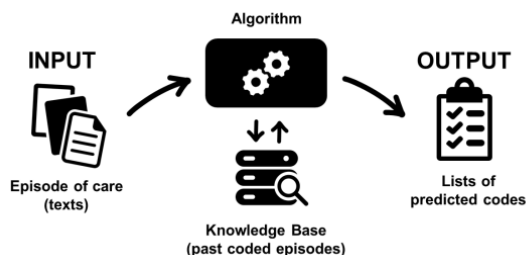


Figure 1. Global workflow of the machine learning encoding. The input is all textual documents from an EOC. The algorithm compares it with past coded episodes contained in the Knowledge Base, and retrieves the k most similar instances, based on statistical textual similarity. From the encoding of these similar episodes, the algorithm infers rankings of candidate CIM and CHOP codes, along with probability scores.

Our study was conducted in the University Hospitals of Geneva (HUG), with a real dataset ranging from 2011 to early 2016 (245 000 EOCs). In Switzerland, EOCs are encoded with two medical classifications derived from the International Classification of Diseases (ICD). Diagnoses are encoded with the CIM classification, issued from ICD-10, German modification; procedures are encoded with the CHOP classification, issued from the ICD-9-CM in 2008 but having evolved since. In Switzerland, both classifications are available in German, French and Italian, but all medical records in Geneva are written in French.

This encoding task is a multi-label automatic text categorization problem [9]. Most reported systems in the literature rely on the popular Support-Vector Machines (SVM). Yet, for large multi-label categorization, SVM has been reported to scale with difficulties [10], requiring the usage of data reduction or feature selection [1]. In our study, with our objectives, we chose to exploit a simple, yet powerful classification algorithm: the k -Nearest Neighbors (k -NN). From a Knowledge Base (KB) containing past encoded instances, this algorithm retrieves the instances that are the most similar to a new instance, and assigns to it the most common codes among these neighbors.

2. Data

At the HUG, encoding is performed by professional coders after an EOC. Coders have limited contacts with the healthcare personnel, but they review all produced documents. There are 4 318 different types of documents, including admission notes, exam or operative reports, progress notes, or discharge letters. Coders have to detect disorders that were addressed during the EOC, and to encode them into CIM codes. For example, for a patient admitted for an alcoholic cirrhosis of liver, the episode will be encoded with the CIM code K70.3 “Cirrhose alcoolique du foie”. The coders also have to detect procedures that were performed. In our example, if a liver biopsy was performed, the coder has to encode it with the CHOP code 50.13.10 “Biopsie du foie transveineuse”. Thus, the whole dataset consists in 245 000 encoded EOCs, from 2011 to early 2016. It is worth noting that what is encoded is EOCs, not documents. In this perspective, all experiments were conducted at the EOC level. For each experiment, a test set of

randomly selected 250 EOCs was submitted to the tool. The algorithm is judged on its capacity to output the codes that were effectively assigned to these episodes by professional coders.

3. Methods

The investigated algorithm is a k-Nearest Neighbors (k-NN). This algorithm performs instance-based learning; it does not learn any discriminative function, but rather exploits a Knowledge Base (KB), containing previously labelled instances, in order to infer real-time decisions [10]. For such text categorization problems with numerous and multiple labels, k-NN has been reported to efficiently scale and to perform comparably with Support Vector Machines [11].

In this study, the KB contains previously encoded EOCs. A similar search engine is designed for retrieving nearest neighbors (i.e. EOCs), based on textual similarity. The Terrier information retrieval platform was used for this purpose [12]. The k parameter sets the number of neighbors exploited for making decisions. Once similar EOCs are retrieved, the second step of the k-NN is to infer encoding for the input episode. For this purpose, the algorithm simply computes what are predominant codes among the retrieved learning instances. The algorithm hence outputs ranked lists of disorders (CIM) and procedures (CHOP) codes, along with confidence scores.

The algorithm is judged on its capacity to output the codes that were effectively assigned to these episodes by professional coders. The evaluation relies on metrics issued from information retrieval [13]. As the tool potentially outputs hundreds of codes – along with decreasing scores – evaluated lists are limited to the top 10 predicted codes. In a semi-automatic workflow, ten is a convenient number of codes easily checkable by the coder. Hence, the recall at rank 10 curves presented in the Results section is the proportion of actual codes that were outputted by the tool in the top 10 ranks. Precision curves are not presented for clarity concerns, but show similar shapes.

4. Results

Preliminary studies (not reported) were conducted with a dataset limited to 2014 EOCs. For both disorders and procedures, the performances of the tool increased with the size of the KB. The most promising observation is that more data leads to better results: for 2014 EOCs, when the size of the KB doubles, the recall is improved by approximately 10%. Thus, with the complete dataset (2011 - 2016), higher performances were expected.

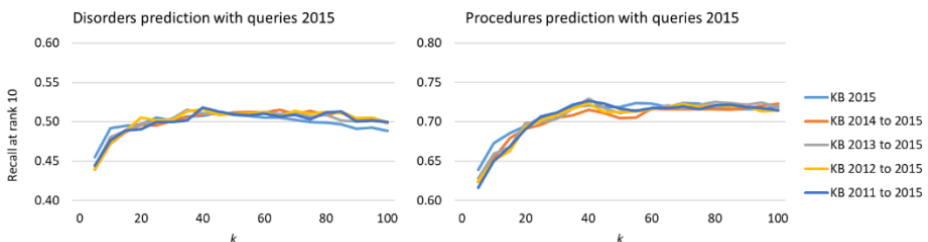


Figure 2. Performances (Recall at rank 10) of the encoding with 2015 queries, for disorders (CIM codes) and procedures (CHOP codes), depending on the k parameter. The KB is populated by previously encoded EOCs, from 52 000 for 2015 only, to 230 000 for the 2011 to 2015 range.

A first set of experiments was conducted for assessing the impact of training data quantity on the model's performances (Figure 2). The test set was a sample of EOCs from 2015, while we progressively increased the number of past EOCs for populating the KB. For both disorders and procedures, all curves are interwoven: the performances of the model do not increase while encoded EOCs from previous years are injected in the KB – yet the past EOCs do not damage the model.

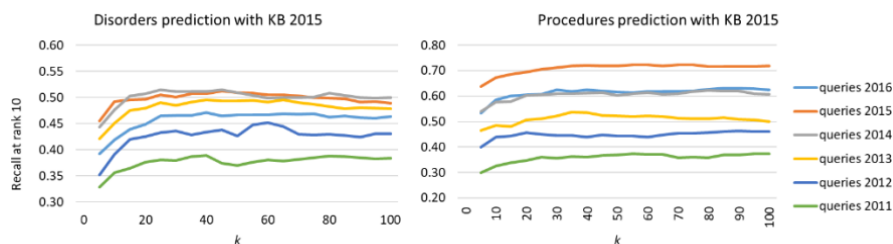


Figure 3. Performances (Recall at rank 10) of the encoding with 2015 KB, depending on the k parameter. Different test sets, containing EOCs sampled in a single year ranging from 2011 to 2016, were used.

A second set of experiments focused on the durability of the model (Figure 3). This time, the KB did not vary and only contained 50 000 encoded episodes from 2015. The model was evaluated with different test sets, each sampled in a single year ranging from 2011 to 2016. The different curves are this time remarkably distinct. With a KB containing previously encoded EOCs from 2015, the best results are reached for input EOCs from 2014 and 2015. Older episodes, from 2013 to 2011, obtain progressively weaker results. These results show that the encoding is considerably year-dependent, and suggest that a model built on a single year is highly perishable – with a potential drop in performances of around -10% per year.

Table 1. Final performances reached for KB and queries from 2015. F-measure is the harmonic mean of precision and recall. This study focused on optimizing recall values.

Coding	F1 at rank 10	P at rank 10	R at rank 10	R at rank 20
Diagnoses	39 %	32 %	51 %	62 %
Procedures	29 %	18 %	72 %	78 %

5. Discussion

The F-measure at rank 10 reached by the tool for 2015 data is 39.3% (see Table 1). This is competitive with F-measures (39.5% [14], 42% [5]) reported in other studies dealing with such large-scale data. Beyond state-of-the-art comparisons, the recall values – 62% for diagnoses and 78% for procedures at rank 20 – must be considered in the light of the inter-annotator agreement, which is a theoretical upper bound. In 2006, this agreement was reported about 79% in the University Hospitals of Geneva [15].

However, the results indicate that the performances of the machine learning model are considerably year-dependent. This could be explained by two main reasons. First, the classifications used for medical encoding in Switzerland evolve over time: every two years for diagnoses, every year for procedures. Codes can be created, merged, or migrated. Second, the coding guidelines proper to each hospital also evolve over time.

Indeed, on the top of diagnoses and procedures codes, rules defining reimbursement are applied by Swiss authorities based on DRGs (Diagnosis Related Groups), unknown from facilities. As two different accurate encodings of the same EOC can lead to differences in the reimbursement, each facility aims at continuously improving its proper coding guidelines, in order to better encode provided healthcare. The consequence is a typical case of concept drift in medical encoding.

6. Conclusion

Are more data better for automatic encoding of EMRs by machine learning? In University Hospitals of Geneva, the answer is mixed. For a unique year, injecting more data in the knowledge base leads to progressively better performances for the algorithm. Yet, injecting encoded episodes from past years does not improve – while does not harm – the performances. Another essential result is that KBs are considerably year-dependent, and a model built on previous years' EOCs appear to be perishable. In this study, a potential drop in performances of around -10% per year has been brought to light. Consequently, great and constant care must be exercised for designing and updating the content of such knowledge bases exploited by machine learning. As Peter Norvig has refined: “More data beats clever algorithms, but better data beats more data”.

References

- [1] R. Kavuluru, A. Rios, Y. Lu, An empirical evaluation of supervised learning approaches in assigning diagnosis codes to electronic medical records, *AI in medicine*, 65(2) (2015), 155-166.
- [2] K.J. O'Malley, K.F. Cook, M.D. Price, K.R. Wildes, J.F. Hurdle, C.M. Ashton, Measuring diagnoses: ICD code accuracy, *Health services research*, 40(5) (2005), 1620-1639.
- [3] Y. Chen, H. Lu, L. Li, Automatic ICD-10 coding algorithm using an improved longest common subsequence based on semantic similarity, *PloS one*, 12(3) (2017), e0173410.
- [4] EMC with Research & Analysis by IDC, The Digital Universe Driving Data Growth in Healthcare, available from: <https://www.emc.com/analyst-report/digital-universe-healthcare-vertical-report-ar.pdf> (2014)
- [5] M. Zeng, M. Li, Z. Fei, Y. Yu, Y. Pan, J. Wang, Automatic ICD-9 coding via deep transfer learning, *Neurocomputing*, 324 (2019), 43-50.
- [6] A. Halevy, P. Norvig, F. Pereira, The unreasonable effectiveness of data. *Intelligent Systems, IEEE*, 24(2) (2009), 8-12.
- [7] M. Banko, E. Brill. Scaling to very very large corpora for natural language disambiguation. In : *Proceedings of the 39th annual meeting on association for computational linguistics*, (2001), 26-33.
- [8] J. Gobeill, E. Pasche, D. Vishnyakova, P. Ruch, Managing the data deluge: data-driven GO category assignment improves while complexity of functional annotation increases, *Database*, (2013) bat041.
- [9] F. Sebastiani, Machine learning in automated text categorization, *ACM computing surveys (CSUR)*, 34(1) (2002), 1-47.
- [10] Z. Yao, W. Ruzzo, A regression-based k nearest neighbor algorithm for gene function prediction from heterogeneous data, *BMC bioinformatics*, 7(1) (2006), 1-11.
- [11] Y. Yang, X. Liu, A re-examination of text categorization methods. *Proceedings of the 22nd annual international SIGIR conference on Research and development in information retrieval*, (1999), 42-49.
- [12] I. Ounis, G. Amati, V. Plachouras, et al, Terrier: A High Performance and Scalable Information Retrieval Platform, *Proceedings of ACM SIGIR'06 Workshop on Open Source Information Retrieval*, (2006).
- [13] C. Manning, P. Raghavan, H. Schütze, Introduction to information retrieval, *Natural Language Engineering*, 16(1) (2010), 100-103.
- [14] A. Perotte, R. Pivovarov, K. Natarajan, N. Weiskopf, F. Wood, N. Elhadad, Diagnosis code assignment: models and evaluation metrics, *Journal of the AMIA*, 21(2) (2013), 231-237.
- [15] P. Ruch, J. Gobeill, I. Tbahriti, A. Geissbühler, From episodes of care to diagnosis codes: automatic text categorization for medico-economic encoding. *AMIA Annual Symposium Proceedings*, (2008), 636-640.

MedCo²: Privacy-Preserving Cohort Exploration and Analysis

David FROELICHER^a, Mickaël MISBACH^a, Juan R. TRONCOSO-PASTORIZA^a,
Jean Louis RAISARO^b, and Jean-Pierre HUBAUX^a

^aEPFL, Switzerland

^bCHUV, Switzerland

Abstract. Medical studies are usually time consuming, cumbersome and extremely costly to perform, and for exploratory research, their results are also difficult to predict a priori. This is particularly the case for rare diseases, for which finding enough patients is difficult and usually requires an international-scale research. In this case, the process can be even more difficult due to the heterogeneity of data-protection regulations, making the data sharing process particularly hard.

In this short paper, we propose MedCo² (pronounced *MedCo square*), a distributed system that streamlines the process of a medical study by bridging and enabling both data discovery and data analysis among multiple databases, while protecting data confidentiality and patients' privacy. MedCo² relies on interactive protocols, homomorphic encryption and differential privacy. It enables the privacy-preserving computations of multiple statistics such as cosine similarity and variance, and the training of machine learning models, on patients that are obliviously selected according to specific criteria among multiple databases.

Keywords. data discovery, data analysis, privacy, confidentiality, homomorphic encryption

1. Introduction

The current trend towards personalized medicine requires researchers to access increasingly larger cohorts of subjects [1], the size of which often exceeds the amount of data available at the researchers' facility. This results in an urgent need for efficient and automatic data sharing mechanisms between medical institutions. Medical data, however, are particularly sensitive, and attempting to share them poses significant threats to the subjects' privacy, as it increases the risk of leaks to ill-intentioned persons or organizations. Medical institutions often refrain from sharing their data due to the increasing number of breaches [2,3], and to their derived reputation and financial implications [4]. Finally, national and international regulations (e.g., HIPAA [5] in the U.S. and the GDPR [6] in the E.U.) impose strong requirements in terms of both confidentiality and restriction of data access that are commonly not met by the existing sharing-solutions [7]. In this environment, in order to access the data, researchers are faced with complex challenges throughout the whole process. Cohort exploration is usually the first step of a medical study; when performed across several medical institutions, it becomes very difficult, as obtaining the necessary legal authorizations and signing agreements between institutions

can take months. This becomes almost impossible when hospitals are in different countries because of the incompatibility of national legal frameworks. Finally, researchers can be left with data of limited value to perform the desired analyses, as they cannot get a priori information (e.g., size, availability and statistical properties such as distribution or moments) to assess the suitability of the data before actually gaining access to them. We propose MedCo², a novel system featuring privacy-preserving cohort exploration and advanced cohort analysis. It enables a researcher to assess the presence of subjects with specific traits and features, compute statistics, and/or train and evaluate machine-learning models on subjects' records, over distributed databases. By relying on distributed and interactive protocols, homomorphic encryption and differential privacy, our system ensures that these queries can be executed yet still protect the confidentiality of data and the privacy of subjects. Finally, MedCo² distributes the trust and avoids single points of failure, as both the storage and the computations are distributed among multiple nodes.

2. System and Threat Models

MedCo² supports a network of mutually distrustful medical institutions that act as data providers (DPs) and hold subjects' records. An authorized researcher (see Figure 1) can run queries without threatening the data confidentiality and subjects' privacy. The DPs are considered semi-honest and the researcher is considered malicious. If required, our system can also use zero-knowledge proofs and an immutable distributed ledger to cope with malicious DPs.

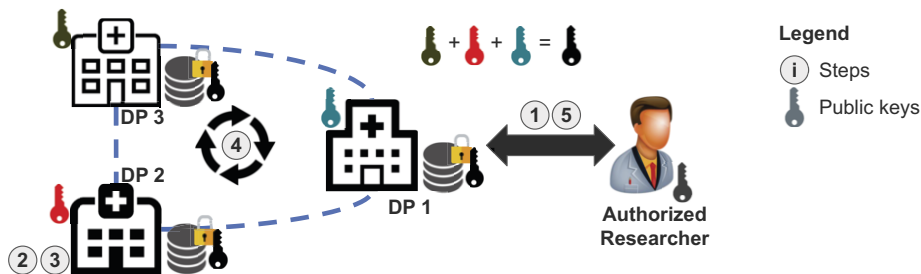


Figure 1. System Model.

3. Proposed Solution

The two main building blocks in MedCo² are MedCo [8] and Drynx [9]. MedCo is an operational distributed system that enables authorized researchers to explore cohorts distributed across several institutions, by filtering subjects with inclusion/exclusion of clinical and genetic criteria. Drynx enables the computation of statistics, such as standard deviation or extrema, and the training and evaluation of machine-learning models on distributed data. In order to ensure data confidentiality and individuals' privacy, both systems rely on distributed interactive protocols, homomorphic encryption and differential privacy. In MedCo², we combine MedCo's capabilities to privately identify a set of sub-

jects matching precise criteria with Drynx's ability to perform computations on the identified subjects' records. As a result, we provide a fully-decentralized system that enables multiple functionalities and enforces privacy and security guarantees that are stronger than in any (to the best of our knowledge) existing solutions [10,11].

We now briefly describe the MedCo² data model, before explaining how a query is executed. DPs have two separate databases: (1) the local database, that is in cleartext and is only accessible from within the institution, and (2) a research database, open to external queries, in which data are encrypted. We define a micro-ETL (extract, transform, load) process that encrypts selected data from the local database and transfers them to the research database. In our system, the data stored in the research databases and all those involved in the query execution are homomorphically encrypted under the DPs' collective key, as described in [9]. A researcher creates a query containing the subjects' selection criteria along with the computation she wishes to perform on the selected subjects' records (step 1 in Figure 1). This query is broadcast to the DPs who privately retrieve the list of subjects satisfying the query and initiate a local data retrieval process (step 2). In this process, each DP stacks the received query and the retrieved subjects' identifiers inside its research database. This stack is periodically emptied by the local operating database that pulls the waiting queries and authorizes them (e.g., based on regulations, local policies). If the query is accepted, each DP performs some query-dependent pre-computation on its local database, before encoding and encrypting its contribution to the secure protocol (step 3). These contributions are then collectively aggregated under homomorphic encryption, to obtain the encrypted results (step 4). At this point, the results are obviously switched from the collective key to the researcher's private key (Collective Key Switching protocol), and sent back to her (step 5). MedCo² executes, in parallel to the query execution, a micro-ETL in which it encrypts the attributes involved in the query and stores them in the research database. This means that when a DP receives another query on the same attributes, it can directly answer from the research database, thus reducing its workload for upcoming queries.

4. Evaluation

We implement and showcase MedCo² within a representative scenario that involves clinical and genetic data sharing. We integrate it in MedCo by enhancing its data model and query language/interpreter to seamlessly enable an authorized researcher to execute operations on a cohort built from MedCo.

We present MedCo² performance for a worst-case scenario: a query that it has not seen before, i.e., we assume that all the computation happens *on the fly*, with no pre-computation. Figure 2 shows the timeline of a query execution in MedCo², comprising the steps of (a) query creation at the client, (b) query broadcast to the DPs, (c) collective query reencryption at the DPs (in order to enable matching queries [8]), (d) local DB execution and result retrieval at each DP, (e) local encoding of the matching patients' data at each DP, (f) collective aggregation of the target function across all DPs, (g) collective key switching of the results to the querier's key, and (h) decrypting and decoding at the querier. In this case, 18132 patients satisfy the filtering criteria and are retrieved among 150,000 patients distributively stored in 12 databases. As shown in [8,13], the overhead time for data discovery (including Step 1, which comprises Query Creation,

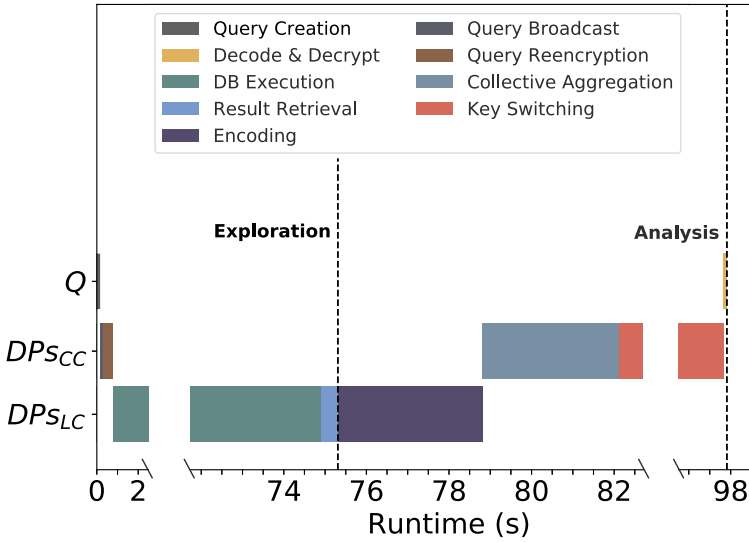


Figure 2. MedCo² Query Execution Timeline. *LC* =Local Computations, *CC* =Collective Computations.

Query Broadcast and Query Reencryption, and Step 2, composed of Database Execution and Result Retrieval) when executed on encrypted data is negligible, in this case 2 seconds, with respect to the same operation executed on cleartext data. Figure 2 also shows that MedCo² can perform a logistic regression on a distributed dataset with 100 features in less than 23 seconds. The distributively trained models achieve accuracy results similar to their centralized, non-secure counterparts, as shown in [9]. We observe that the whole MedCo² process is performed in less than 98 seconds if 18132 patients out of 150000 satisfy the selection criteria. The *DPs*' local query execution in MedCo (DB execution in Figure 2) is the most expensive operation, as it linearly depends on the number of selection criterias in the query and on the total number of observations in the database (each patient has thousands of observations).

Finally, MedCo² enables a researcher to train and evaluate linear and logistic regression models on a chosen cohort without accessing the actual subjects' records. This makes the task more cost-effective and streamlines a process that is at the core of many common medical workflows such as GWAS.

5. Conclusion

MedCo² addresses the privacy and security challenges of highly distributed medical-data sharing networks by avoiding single points of failure and by providing the following unparalleled functionalities and properties: (a) privacy-preserving cohort exploration and analysis over distributed databases held by distrustful *DPs*, (b) distributed training of simple machine learning models, (c) data confidentiality and individuals' privacy through collective keys and distributed protocols, and (d) a usable example scenario through its integration in a deployed system featuring a modern graphical user interface [8]. MedCo² aims at enabling the promises of personalized health by facilitating and streamlining the data discovery and analysis processes.

Acknowledgements

This work was supported in part by the grant #2017-201 (DPPH) of the Swiss strategic focus area Personalized Health and Related Technologies (PHRT), and by the grant #2018-522 (MedCo) of the PHRT and the Swiss Personalized Health Network (SPHN).

References

- [1] P. Mahon and J.M. Tenenbaum, Paths to Precision Medicine - A Perspective., *Journal of Precision Medicine*, 2015.
- [2] Ponemon Institute. Sixth annual benchmark study on privacy security of healthcare data, *Tech. report*, 2016.
- [3] Healthcare Data Breach Statistics, <https://www.hipaajournal.com/healthcare-data-breach-statistics/>, Accessed: July 03, 2019.
- [4] A huge trove of medical records [...] found exposed, <https://tcrn.ch/2Cp6hzv>, Accessed: July 15, 2019.
- [5] The health insurance portability and accountability act (hipaa), <https://www.hhs.gov/hipaa/index.html>, Accessed: June 26, 2018.
- [6] EU Parliament. The EU General Data Protection Regulation (GDPR). <http://www.eugdpr.org/>. Accessed: July 15, 2019.
- [7] SHRINE, <https://www.i2b2.org/work/shrine.html>, Accessed: June 29, 2018.
- [8] J.L. Raisaro, J. R. Troncoso-Pastoriza, M. Misbach, J. A. Sá Sousa, S. Pradervand, E. Missiaglia, O. Michielin, B. A. Ford, and J.-P. Hubaux, Medco: Enabling privacy-conscious exploration of distributed clinical and genomic data. *IEEE/ACM TCBB*, 2018.
- [9] D. Froelicher, J. R. Troncoso-Pastoriza, J. A. Gomes de Sá E Sousa and J.-P. Hubaux, Drynx: Decentralized, Secure, Verifiable System for Statistical Queries and Machine Learning on Distributed Datasets. <http://arxiv.org/abs/1902.03785>, 2018.
- [10] H. Cho, D. J. Wu and B. Berger. Secure genome-wide association analysis using multiparty computation, *Nature biotechnology*, 2018.
- [11] K. A. Jagadeesh, D. J. Wu, J. A. Birgmeier, D. Boneh and G. Bejerano. Deriving genomic diagnoses without revealing patient genomes. *Science*, 357(6352), 692-695, 2017.
- [12] i2b2, <https://www.i2b2.org/>, Accessed: June 29, 2018.
- [13] D. Grishin, J. L. Raisaro, J. R. Troncoso-Pastoriza, K. Obbad, K. Quinn, M. Misbach, J. Gollhardt, J. Sa, J. Fellay, G. M. Church and J.-P. Hubaux, Citizen-Centered, Auditable, and Privacy-Preserving Population Genomics, *bioRxiv*, <https://www.biorxiv.org/content/10.1101/799999v1>.

Mixed and Augmented Reality Tools in the Medical Anatomy Curriculum

Melanie ROMAND^{a,1}, Daniel DUGAS^b

Christophe GAUDET-BLAVIGNAC, BSC CS, MMED^a Jessica ROCHAT^a and
Christian LOVIS, MD, MPH, FACMI^a

^a*Division of Medical Information Sciences, University Hospitals of Geneva, Geneva, Switzerland*

^b*Autonomous Systems Lab, ETH, Zurich*

Abstract. The use of augmented reality (AR) in the medical field has grown since 2007-2013 and was first introduced in surgical departments. AR and mixed reality (MR) allow us to explore complex structures, observe phenomena that are difficult or impossible to see otherwise, and interact with the virtual structures they display. Recently, they are beginning to be adopted for education. This work examines whether new AR and MR technological tools, when used in the context of anatomical teaching, can allow for strengthening of teaching quality - for example by overcoming new or existing constraints such as the limited availability of dissection specimen. This work also considers how these technologies are to be applied efficiently and practically in teaching. An attempt is made to answer these questions in three stages: i) a non-systematic review of the literature ii) a review of augmented reality solutions for anatomy available on four different platforms iii) a 30-person study of the usability of augmented reality. The results show that there is potential for AR and MR to supplement anatomical teaching, but that traditional methods remain indispensable.

Keywords. augmented reality, mixed reality, anatomical teaching, e-learning, digital tools for supplementing medical teaching

1. Introduction

Anatomy is a fundamental branch of the medical curriculum. As the human body is the target of daily interventions and investigations, this discipline is a cornerstone of physicians' medical skills. Anatomical knowledge allows for safe and quality medical practice, through the understanding of decisions and actions taken and the fact that it contributes to effective communication between health professionals [1]. Given its importance, its teaching should be as effective as possible, in terms of accuracy of the acquired knowledge, and its resilience.

Since the Renaissance, the study of corpses and their dissection have been the main means of teaching anatomy [2]. Today, this discipline is traditionally taught in the pre-clinical years and provides basic knowledge in the four fields of macroscopic anatomy

¹Corresponding Author: Melanie Romand, Université de Genève, 24 rue du Général-Dufour 1211 Genève, Switzerland; E-mail: Melanie.Romand@etu.unige.ch

(including dissection), neuroanatomy, histology and embryology. Course material generally includes slideshows with verbal explanations, bodies for dissection, presentations of already dissected body parts and books with two-dimensional images [3].

Since its inception, the field of anatomical teaching faces several practical constraints. These include the time allocated within the medical curriculum, the practicality of using human bodies and the ethical and emotional dimension surrounding the handling of corpses by students. Some of those constraints are relatively new, or are currently increasing in severity. The evolution of the constraints which apply to the field of anatomical teaching, raises the question of whether the tried and true methods are still optimal. In addition, technological advances make previously unavailable medias and modalities possible. This raises the questions of how modern technologies could be used to supplement the teaching of this discipline, whether applying these tools would lead to improved efficiency in practice, and what additional challenges technologically augmented anatomical teaching might present. Previous work [3] shows that anatomy learning is most effective when the desired structures can be examined from all angles and at any time. Recent technological advances in the field of virtual, mixed and augmented reality have led to the emergence of widely available consumer software, academic or commercial, which focus on anatomy learning.

One aspect of this work consisted in evaluating the already-available tools and solutions in the academic and commercial field, through a literature review and application inventory. In a second stage, two of the most promising mixed-reality applications were selected in order to conduct a usability study, with the participation of medical students. This usability study focused on the hands-on evaluation of two applications available on Microsoft's mixed reality headset, "Hololens". The results collected throughout this study are evaluated in order to extract recommendations in proposing an evolution in the teaching of this discipline.

2. Methodology

2.1. *Nonsystematic Literature Review*

A literature search was conducted on the following search engines and knowledge bases: PubMed, Google scholar, Embase/Elsevier/ScienceDirect, Wiley Online Library, Web of science, and more. The inclusion criteria for literature was: The article must discuss the topic of AR/MR in the general field of medicine OR the article must discuss the topic of AR/MR in the teaching of anatomy.

2.2. *Study of Available Solutions*

Presently, the most widely used AR devices are consumer smartphones and tablets [4]. In the pursuit of studying available AR and MR solutions, the three most promising AR and MR platforms were identified, and software available on those platforms was tested and evaluated. Selection criteria: Software was selected for evaluation according to the following criteria: i) the software must focus on human anatomy ii) the software must present significant AR or MR capabilities. iii) the software must be available for use on the Hololens, Apple device, or Google Play device platforms. Of the selected software, a

sub-selection was made based on available user-review ratings. Finally, a small portion of the selected software was excluded due to practical limitations. Each remaining software solution was tested in-person, and evaluated according to a consistent evaluation scheme. The evaluation scheme was formed to include as much potentially relevant information as possible for anatomical teaching providers.

2.3. Usability Study

The usability study conducted in the course of this work can be modeled in three parts. The experiment set-up: the target demographic for the study was medical students, in the 2nd to 5th year. 1st and 6th year students were not targeted due to a lack of complete ongoing anatomical curriculum. 30 medical students signed-up for and participated in the study. The hands-on experiment: two solutions were selected based on their ease-of-use, relevance and completeness: "Dynamic Anatomy" and "Holoanatomy", both MR applications intended for and tested on the HoloLens. Hands-on experimentation: during 30 minutes, participants were to interact with both selected MR solutions, in a semi-supervised manner. Structure was given, but not imposed in the participants' interaction with each application. Assistance was available and given on request. Then a usability questionnaire, developed based on the standard set by [5], was filled in by participants after they had completed all other tasks. The questionnaire contained 4 sections: 1) **statements** concerning their experience with MR, 2) perceived **positive and negative** aspects, 3) a **multiple-choice** question regarding their preferred mode of anatomy teaching, and 4) **arguments** supporting their choice on the multiple-choice question.

3. Results

3.1. Nonsystematic Literature Review

Using the criteria outlined in Subsection 2.1, 27 existing articles were selected for review. 8 of the reviewed articles relate to describing evolution in the topic of AR in anatomical teaching. 19 of the reviewed articles concern tools presently available to apply AR and MR technologies.

Work by [6] describes the evolution of AR use in medical scenarios, and identifies a rise in activity for these topics in the years 2007-2013. The terms MR and AR are used inconsistently in the literature. As a result, tracing back the evolution for MR specifically is challenging. Technical limitations have contributed to hindering the introduction of AR to the medical field [7] Despite this, research projects involving the use of AR in medical scenarios, and specifically in anatomical teaching were identified. Previous works look at the use of AR in surgery [6], neurosurgery [8], and oto-rhino-laryngology [9] Research towards the use of AR for anatomical teaching dates back at least as far as 1997 [10]. Since then, several works [11] have focused on applying AR in medical teaching.

3.2. Available AR and MR Solutions

52 applications were selected, 26 of those available on the Apple ecosystem, 16 on the Google Play ecosystem, and 10 on the HoloLens. Of those, 35 applications (respectively 17, 8, and 10 for Apple, Google Play, and HoloLens) could be tested in person.

Table 1.: Open-ended 3-positive-3-negative-aspects questionnaire results

Hololens for anatomy teaching				MR for anatomy teaching			
Positive	%	Negative	%	Positive	%	Negative	%
Ease of use	26%	Discomfort	35%	3D Visualization	22%	Discomfort	30%
Adjustability	10%	Field of view	22%	Practicality	18%	Missing touch	17%
Convenience	9%	Price	19%	Super-real		Space required	17%
Comfort	9%	Software bugs	10%	visualization	15%	Little benefit	
Low weight	9%	Availability	7%	Stimulating	12%	compared to PC	10%
Interactivity	9%	Non-trivial		Interactivity	10%	Lack of realism	8%
Portability	7%	learning curve	7%	Increased		No social aspect	8%
Intuitiveness	7%			understanding	9%	Lack of real body	5%
Image quality	7%			Innovative	8%	Non-mature tech.	5%
Voice, gesture				Add, remove, and			
recognition	7%			isolate structures	6%		

The overall resulting impression is as follows: At the time of this study, MR/AR applications exist which could be used as tools to supplement anatomy teaching. However commercial solutions do not yet exist for every body part, or are not suited for all specific teaching modalities. A knowledge base of these evaluations was compiled and is made freely available (www.dugas.ch/public/realite-mixte-app-evaluation-data). Based on these evaluations, the Hololens platform was selected for the usability study described in Subsection 2.3.

3.3. Usability Study

The usability questionnaire contained 4 sections. The results obtained for each section are as follows:

The **statements** supplied in the questionnaire were graded on a scale from 1 (worse) to 7 (best). The mean score by category was 5.03 for the usefulness, 5.42 for ease of use, 6.2 for ease of learning and 5.63 for satisfaction. The overall grade was 5.47. Full results are available on request.

Responses to the open-ended **3-positive-3-negative-aspects** questionnaire were grouped into conceptual categories, semantics aside. Those categories, and the frequency at which participant responses fell into the relevant category are shown in Table 1.

As shown in Table 2, in the concluding **multiple-choice** poll, following hands-on testing of MR tools and existing software, an overwhelming majority of students expressed a desire for both classical methods and technological tools to be used in the teaching of anatomy in the medical curriculum.

In the open-ended supporting **arguments** question, participants were asked to motivate their choice. Responses to this open-ended question were grouped based on conceptual similarity. Arguments for using MR in anatomical teaching fell into the following categories: i) Access to specimens, and in particular the ability for home-work, which is difficult with real specimens. ii) Access to otherwise-impossible (referred to as *super-real* in this work) visualizations, such as dynamical properties, hidden structures, etc. Arguments for traditional methods fell into the following categories: i) realism, and the ability to show details which digital media can not practically capture ii) the psychological impact of real bodies on professional maturity iii) the importance of all senses in anatomy learning.

Table 2.: Preferred modality to be used in anatomical courses

Proposed general statement	selected %
Only MR methods, with Hololens	3%
Only currently used, traditional methods	11%
A combination of traditional and MR methods	86%

4. Conclusion

It has been shown that practical factors such as a high number of students, passive participation, difficult concepts [12], and a large quantity of material to be learned, can lead to sub-optimal or superficial learning for students in medical anatomy. Proper spacing of study sessions is important for optimal long-term memorization [13]. Use of AR and MR anatomical teaching could pose several advantages, such as the possibility of increasing the frequency of learning instances, interaction with models and specimens which are rare in practice, and super-real visualizations. This work shows that medical students are open to the introduction of technological tools to supplement, though not replace, traditional anatomy teaching. Potential future work involves extending the scope of the present study, and conducting a feasibility study from the perspective of anatomy teachers.

References

[1] B. W. Turney, "Anatomy in a modern medical curriculum," *Annals of the Royal College of Surgeons of England*, vol. 89, no. 2, pp. 104–107, 2007.

[2] J. C. McLachlan, J. Bligh, P. Bradley, and J. Searle, "Teaching anatomy without cadavers," 2004.

[3] C. Moro, Z. Štromberga, A. Raikos, and A. Stirling, "The effectiveness of virtual and augmented reality in health sciences and medical anatomy," *Anatomical Sciences Education*, vol. 10, pp. 549–559, nov 2017.

[4] C. Kamphuis, E. Barsom, M. Schijven, and N. Christoph, "Augmented reality in medical education?," *Perspectives on Medical Education*, vol. 3, pp. 300–311, sep 2014.

[5] A. Lund, "(PDF) Measuring Usability with the USE Questionnaire."

[6] M. Eckert, J. S. Volmerg, and C. M. Friedrich, "Review of Augmented Reality in Medicine (Preprint)," *JMIR mHealth and uHealth*, vol. 7, p. e10967, apr 2019.

[7] J. Soeiro, A. P. Claudio, M. B. Carmo, and H. A. Ferreira, "Visualizing the brain on a mixed reality smartphone application," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2015-Novem, pp. 5090–5093, IEEE, aug 2015.

[8] C. Lee and G. K. C. Wong, "Virtual reality and augmented reality in the management of intracranial tumors: A review," *Journal of Clinical Neuroscience*, vol. 62, pp. 14–20, apr 2019.

[9] J. L. McJunkin, P. Jiramongkolchai, W. Chung, M. Southworth, N. Durakovic, C. A. Buchman, and J. R. Silva, "Development of a Mixed Reality Platform for Lateral Skull Base Anatomy," *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otolaryngology and Neurotology*, vol. 39, no. 10, pp. e1137–e1142, 2018.

[10] J. P. Rolland, D. L. Wright, and A. R. Kancherla, "Towards a novel augmented-reality tool to visualize dynamic 3-D anatomy," in *Studies in Health Technology and Informatics*, vol. 39, pp. 337–348, 1997.

[11] S. Küçük, S. Kapakin, and Y. Göktas, "Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load," *Anatomical sciences education*, vol. 9, no. 5, pp. 411–421, 2016.

[12] R. Barmaki, K. Yu, R. Pearlman, R. Shingles, F. Bork, G. M. Osgood, and N. Navab, "Enhancement of Anatomical Education Using Augmented Reality: An Empirical Study of Body Painting," *Anatomical Sciences Education*, feb 2019.

[13] L. Lochner, H. Wieser, S. Waldböth, and M. Mischo-Kelling, "Combining traditional anatomy lectures with e-learning activities: how do students perceive their learning experience?," *International journal of medical education*, vol. 7, pp. 69–74, 2016.

Modulation of Medical Condition Likelihood by Patient History Similarity

Jonathan TURNER^{a,1}, Dymphna O’SULLIVAN^b and Jon BIRD^c

^a City, University of London, London, UK

^b Technological University Dublin, Dublin, Ireland

^c University of Bristol, Bristol, UK

Abstract. Introduction: We describe an analysis that modulates the simple population prevalence derived likelihood of a particular condition occurring in an individual by matching the individual with other individuals with similar clinical histories and determining the prevalence of the condition within the matched group. **Methods:** We have taken clinical event codes and dates from anonymised longitudinal primary care records for 25,979 patients with 749,053 recorded clinical events. Using a nearest neighbour approach, for each patient, the likelihood of a condition occurring was adjusted from the population prevalence to the prevalence of the condition within those patients with the closest matching clinical history. **Results:** For conditions investigated, the nearest method performed well in comparison with standard logistic regression. **Conclusions:** Results indicate that it may be possible to use histories to identify ‘similar’ patients and thus to modulate future likelihoods of a condition occurring.

Keywords. EHR; machine learning; clinical terminologies

1. Introduction

The advent of electronic information systems in recent decades has enabled a rapid increase in data creation and collection in many disciplines, not least in healthcare, where we now have large amounts of data stored in individuals’ electronic health records. These data have been acquired during the care of individuals with a variety of aims: to improve the care of the individuals; to improve communications between individuals’ care givers; to maintain records of care; and to enable appropriate charging to payers [1]. Electronic records generally now hold data dating back 30 years or more and exist in various healthcare enterprises and departments for a variety of purposes – e.g. booking systems, GP systems, lab test results, imaging – and contain a number of data items, such as demographic information, test results, diagnoses, treatments, clinic appointments and charging information. In the work presented here, a set of longitudinal primary care records was analyzed to see if patients with similar recorded histories of symptoms and diagnoses were likely to share similar future diagnoses. The analysis of Electronic Health Records (EHR) data for risk prediction is an active area of research [2] however scant work using solely primary care records data for general future conditions likelihood was found. Miotto et al [3] used secondary care EHR data to make such predictions for a number of conditions using deep learning, achieving an

¹ Corresponding author: Jonathan Turner, Centre for Health Informatics, City, University of London, Northampton Square, London EC1V 0HB, United Kingdom; E-mail: jonathan.turner@city.ac.uk.

F-score of 0.236 for their predictions. Weng et al [4] tested four machine learning algorithms (random forest, logistic regression, gradient boosting machines, neural networks) to predict cardiovascular risk. McCormick et al [5] describe a Bayesian hierarchical model for the selection of association rules, predicting future medical histories for individuals on the basis of common clinical histories. Wang et al [6] developed a multilinear sparse logistic regression for risk prediction with patient EHRs for joint risk prediction (two or more diseases that are related to each other in terms of sharing common comorbidities, symptoms, risk factors). In summary, other work to date has primarily been on secondary care records for patients being treated in hospital settings as part of single episodes; the work we describe here investigates the use of primary care records that record events over many years.

2. Objective

With many symptoms being common to multiple diseases, there is a challenge in producing an initial diagnosis or recommendation for diagnostic tests from a set of symptoms that could have been produced by a number of diseases. Often the initial choice of diagnosis or testing is based on a clinician's impression of the likelihood of that condition in a general population; however the opportunity may exist for modification of these likelihoods based on individuals' recorded histories. In this work we aim to investigate the potential of applying machine learning methods to medical histories, in order to see if we can predict whether an individual's chances of being diagnosed with a particular condition are modulated from general population prevalence by comparison of their previous diagnosis and symptom histories with others' diagnosis and symptom histories. Jenssen and Kenyon note that 'medical decision-making is about assessing risk and weighing competing probabilities in the setting of imperfect information' [7]. We used coded histories from primary care records as the basis for this work. Primary care records, in general, utilize coding systems to store diagnosis and other information about patients. Common coding systems used for these records include ICD9, the dominant system in primary care in the USA and now being replaced by ICD10; and the Read Code system, the dominant system in the UK but now being replaced by SNOMED CT. Data used in this study were derived from UK patients, coded using Version 3 of the Read Codes (CTV3), and US patients, coded using ICD9. We chose CTV3 as the primary coding system. Some data required translation from ICD9 to CTV3 using a look-up table created from existing look-up tables for ICD9 to SNOMED CT from NIH [8] and for SNOMED CT to Read CTV3 from NHS Technology Reference Data Update Distribution Service [9]. Each CTV3 term is associated with a single concept and consists of a 5 byte code with an optional 2 byte. Codes comprise letters, numbers and trailing decimal points and fall into one of three categories: findings and diagnoses; processes; and medications. All codes are arranged in a hierarchy with a table listing all parent-child relations.

Figure 1 shows a small extract from the



Figure 1. An example of codes in the CTV3 hierarchy.

CTV3 hierarchy, illustrating the relationship between codes relating to bacterial chest infections. This Figure shows that Acute haemophilus influenza bronchitis is a Haemophilus influenzae infection; there are four other codes that are child codes of Haemophilus influenzae infection. Haemophilus influenzae infection itself is a Haemophilus infection, which in turn is a Bacterial disease. Bacterial disease is an Infective disorder, a child code of Clinical findings, a category coming under the root node of the CTV3 tree.

3. Methods

The data sources were sets of longitudinal clinical records derived from general practice data across the UK and USA, de-identified prior to being made available to us for this work. Data were presented as a set of records of events, including patient age, gender, and codes of events including diagnoses, symptoms, administrative events and prescriptions, together with the year of the event and history of smoking and alcohol consumption. There were 25,979 patients in the full data set (49.3%M, 50.7%F), with a total of 749,053 recorded items, of which 375,312 items were diagnostic events. Age distribution and prevalence of common conditions (as determined by [10]) in the full data set was not different to that present in the general population. For this work, only diagnosis and symptom codes and those codes giving information about smoking and alcohol consumption were retained from the records in the data set. The Konstanz Information Miner (KNIME) [11] was used to prepare the data and merge the data sets; statistical analysis was performed using R [12]. All software was run on an Asus N56VM laptop with Intel Core i7-3610QM 2.30 GHz CPU with 8GB RAM running 64-bit Windows 10.

For each record, the codes retained were used to determine its k nearest neighbour (KNN) records by use of the binomial method [13] using a program written in R for this work. Optimum values for k , the number of nearest neighbours, were established for each condition by experimentation on a training set of records removed from the complete data set, where the F-measure of accuracy was optimized over a range of values for k . The most effective depth in the CTV3 tree at which to aggregate codes from deeper levels was similarly determined by experiment. Calculations were repeated for a set of medical conditions. The process was run separately for each condition, with the condition of interest hidden from the information used for the distance calculation. For each record, the prevalence of the condition in the KNN was compared to the prevalence in the remainder of the data set; those records where the prevalence in the neighbours was significantly ($p < 0.05$) higher than in the remaining records, were assigned a prediction of positive for the condition of interest, otherwise the prediction was negative. Predictions were then compared to the actual presence in each record.

4. Results

An example is given for a single condition, allergic rhinitis. A Fagan nomogram [14] of the changed probability of an individual having the condition is illustrated in Figure 2. The prior probability of an individual having the condition, i.e. the population prevalence, is 0.178. After implementing the nearest neighbours algorithm, (the 'test') those predicted to have the condition had a probability of having the condition of 0.252.

The sensitivity of the test was 0.599; the specificity was 0.614. Results for six conditions are shown in Table 1. The results include comparison with a simple logistic regression (LR) method, chosen as an established prediction method [2].

KNN generally performed better than LR when judged by F-score, sensitivity and specificity, for all conditions. Low prevalence conditions performed poorly with LR, predicting no positive cases. Because of LR’s better performance in positive likelihood ratio, it also gave higher results for odds ratio than KNN, suggesting that if LR predicts positive for a condition then there is greater confidence in the record containing that condition than if KNN predicts positive. However, if LR predicts negative for a record then there is less confidence that the record does truly not contain the condition than if the KNN predicted negative. F1 scores achieved for the more common conditions compared well with those achieved by Miotto et al [3]. Additionally, the KNN method presents an explainable method that may be more convincing than other machine learning methods in a clinical setting [4].

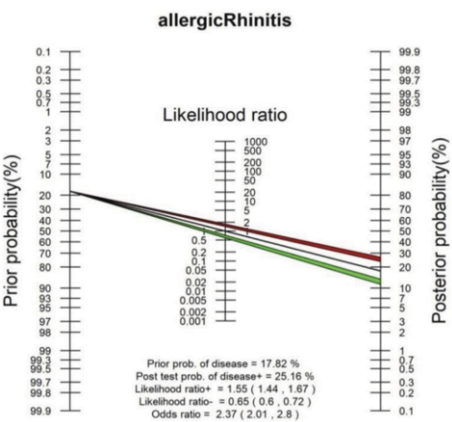


Figure 2. Results for allergic rhinitis using k-nearest neighbours.

Table 1. Comparison of KNN and LR prediction methods for six conditions.

Condition	Number in data set	Method	F1 Score	F2 Score	Sensitivity	Specificity	Positive Likelihood Ratio	Negative Likelihood Ratio	Odds Ratio
Allergic Rhinitis	2297	KNN	0.354	0.469	0.599	0.614	1.550	0.654	2.37
		LR	0.113	0.080	0.067	0.974	2.518	0.959	2.626
Bronchitis	2004	KNN	0.328	0.433	0.552	0.661	1.625	0.679	2.394
		LR	0.139	0.981	0.223	0.164	7.240	0.881	5.146
Obesity	2000	KNN	0.334	0.443	0.566	0.670	1.713	0.648	2.643
		LR	0.409	0.326	0.287	0.979	13.422	0.729	18.420
Gastroparesis	26	KNN	0.017	0.039	0.286	0.944	5.06	0.757	6.684
		LR	0	0	0	0.999	0	0	0
Autism	25	KNN	0.016	0.034	0.167	0.970	5.493	0.854	6.392
		LR	0	0	0	1	0	0	0
Eczema	670	KNN	0.229	0.369	0.623	0.791	2.984	0.476	6.265
		LR	0.137	0.095	0.079	0.996	21.649	0.925	23.413

5. Conclusions and Future Work

The results of our analysis indicate that it is possible to apply nearest neighbour techniques using medical histories to modulate future likelihoods of a condition occurring in individual patients. The technique employed in this work showed

performance improvements over logistic regression methods particularly for conditions of low prevalence, although there were limitations due to the relatively small data set size. This method is suggested as a useful low-cost screening tool. Early discovery of a condition can often improve the chances of successful treatment or of mitigation of the effects of the condition and many conditions can be detected through screening tests on asymptomatic individuals. Disadvantages of conventional screening tests include risks of harm to the individual if the screening test is invasive; inconvenience and/or stress to the individual; and financial cost to the individual and/or the health care system. It is therefore advantageous to perform screening tests on subsets of the population selected to be at great risk of the condition being screened for [15].

We intend to extend our analysis in a number in a number of areas. We will repeat the analysis using the existing methods but translating event codes to SNOMED CT and comparing results to those described here. We intend to optimize the methods used to address the slow run time experienced in this analysis. Run time has implications for use in clinical practice: currently the system takes around 30 minutes to run on a sample of 5,000 records. This is likely to be too long for opportunistic likelihood predictions for an individual arriving in a clinic, although acceptable for use in identification of individuals appropriate for screening or invitation to consultation.

References

- [1] Coiera E. *Guide to Health Informatics*. CRC Press, London, 2015.
- [2] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calsterad B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology* **110** (2019): 12-22.
- [3] Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Scientific Reports* **17** (2016), 6:26094.
- [4] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can Machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE* **12** (2017), e0174944.
- [5] McCormick T, Rudin C, Madigan D. Bayesian Hierarchical Rule Modeling For Predicting Medical Conditions. *Annals of Applied Statistics*, **6** (2012), 652-668.
- [6] Wang X, Wang F, Hu J, Sorrentino R. Exploring joint disease risk prediction. Proceedings of the Annual Symposium of American Medical Informatics Association (AMIA), 2014.
- [7] Jessen BP, Kenyon CC. Probability, Uncertainty, and Value in Inpatient Diagnosis: Connecting the Dots. *Hospital Pediatrics*, **5** (2015) 403-405
- [8] US National Library of Medicine. *ICD-9-CM Diagnostic Codes to SNOMED CT Map, Version 2016* (2016). Available at: https://www.nlm.nih.gov/research/umls/mapping_projects/icd9cm_to_snomedct.html (accessed 3rd January 2018).
- [9] Health and Social Care Information Centre. *NHS Data Migration Releases*. (2013). Available at <https://isd.digital.nhs.uk/trud3/user/guest/group/0/pack/9/subpack/9/releases> (accessed 13th January 2020).
- [10] Practice Fusion. *The 25 Most Common Medical Conditions*., 2016. Available at <https://www.practicefusion.com/blog/25-most-common-diagnoses/> (accessed 13th January 2020).
- [11] Berthold MR, Cebon N, Dill F, Gabriel TR, Kotter T, Meini T, et al. KNIME: The Konstanz Information Miner. In Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds): *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*. Berlin: Springer; 2008. p. 26-31.
- [12] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org/> (accessed 10th October 2018).
- [13] Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. *BMC Medical Informatics and Decision Making* **16** (2016), 197-208.
- [14] Fagan TJ. Nomogram for Bayes theorem. *New England Journal of Medicine* **293** (1975), 257.
- [15] Hobbs F. Cardiovascular disease: different strategies for primary and secondary prevention? *Heart* **90** (2010), 1217-1223.

MQT-Tz: Secure MQTT Broker for Biomedical Signal Processing on the Edge

Carlos SEGARRA ^{a,b,1}, Ricard DELGADO-GONZALO ^a and Valerio SCHIAVONI ^b

^a CSEM, Neuchâtel, Switzerland

^b Université de Neuchâtel, Neuchâtel, Switzerland

Abstract. Physical health records belong to healthcare providers, but the information contained within belongs to each patient. In an increasing manner, more health-related data is being acquired by wearables and other IoT devices following the ever-increasing trend of the *Quantified Self*. Even though data protection regulations (e.g., GDPR) encourage the usage of privacy-preserving processing techniques, most of the current IoT infrastructure was not originally conceived for such purposes. One of the most used communication protocols, MQTT, is a lightweight publish-subscribe protocol commonly used in the Edge and IoT applications. In MQTT, the broker must process data on clear text, hence exposing a large attack surface for a malicious agent to steal/tamper with this health-related data. In this paper, we introduce MQT-Tz, a secure MQTT broker leveraging Arm TRUSTZONE, a popular Trusted Execution Environment (TEE). We define a mutual TLS-based handshake and a two-layer encryption for end-to-end security using the TEE as a trusted proxy. We provide quantitative evaluation of our open-source PoC on streaming ECGs in real time and highlight the trade-offs.

Keywords. wearables, mHealth, secure broker, MQTT, mosquito, TrustZone

1. Introduction

Personalized health and medicine has the potential of being the next revolution in healthcare. It is also referred as the P4 medicine (Predictive, Preventive, Personalized, and Participatory), and provides the opportunity to benefit from more targeted and effective diagnoses and treatments [1]. One of the driving forces behind this tendency is the increasing medicalization of wearable technology [2] and mobile health (mHealth) apps [3]. In order to enable these technologies, complex processing IoT pipelines are gradually being deployed or repurposed. When the data-in-motion are vital signs, protecting user's privacy becomes a topic of crucial importance. Recent data protection regulations (e.g., GDPR [4]) stress the importance of protecting sensitive information against malicious attackers or untrusted cloud providers.

Message Queuing Telemetry Transport (MQTT) [5] is one of the most commonly-used communication protocols in IoT. In spite of that, it is not included in some of the most extended Medical-Grade data exchange standards [6,7]. It follows a publish-subscribe architecture specially designed for environments with limited memory and reduced network bandwidth. In such circumstances, MQTT has proven to be more adapted to the IoT than classical protocols such as HTTP [8]. In MQTT, a *client* publishes data to a *topic*

¹ Corresponding Author: Carlos Segarra; E-mail: carlossegarragonzalez@gmail.com.

and the *broker* forwards it to each client that previously subscribed to it. The protocol is currently used in a variety of settings: data generation by sensors, pre-processing on the edge, and forwarding to the cloud. Examples include live heart-rate data [9,10], smart-grids [11], or building management systems [12]. Most MQTT implementations support TLS for transport security in the client-broker link, preventing malicious actors from spoofing application data. However, the broker itself still exposes a great attack surface [13].

In order to protect the privacy of health-related data, we present MQT-TZ, a secure implementation of the MQTT broker leveraging Arm TRUSTZONE, a Trusted Execution Environments (TEE), widely available on edge devices [14]. TRUSTZONE is a security feature available in recent Arm processors that enables system-wide hardware isolation for trusted software [15]. Our prototype builds atop *mosquitto* (<https://mosquitto.org>), a popular MQTT broker implementation, and includes persistent storage of client's keys in Arm's tamper-proof secure storage, as well as TEE-protected re-encryption of application data. These security enhancements make our ecosystem compliant with the "Services Secure Interface" [6] described by the Personal Connected Health Alliance, and address several attack vectors listed [7] by the IHE. We also consider linking our secure broker to a larger storage utility where data-at-rest is encrypted and its origin authenticated by MQT-TZ.

The paper is organized as follows. In Section 2, we describe the technical architecture and implementation of MQT-TZ. Then, in Section 3, we evaluate its performance and robustness at processing 1-lead ECGs in real time. Finally, in Section 4 we expose our main conclusions and propose further lines of research.

2. MQT-TZ: Securing the MQTT Broker

2.1 Architecture & Component Description

TRUSTZONE splits the system in a hardware-protected trusted part (the TEE) and an untrusted one (also called Rich Execution Environment, or REE). We add an encryption layer in MQTT's payload using client-specific keys stored in Arm's secure storage. This way, application data is only processed in clear inside the TEE. For the additional key-provisioning, we redefine the client authentication in the mutual TLS handshake to prevent the REE from gaining access to clients' keys.

The overall workflow looks as follows. Data travels two-fold encrypted from the client to the broker (Fig. 1: part 1). Once the client access is confirmed, Fig. 1: part 2, the subscribers for the given topic are retrieved and the payload forwarded (Fig. 1: part 3). Then, encrypted data is transferred to the TEE (Fig. 1: part 4). The origin and destination client keys are retrieved (fig. 1: part 5-7), the payload is re-encrypted, and sent back to the REE (Fig. 1: part 8) and to the subscriber (Fig. 1: part 9).

Two-Step Handshake. MQT-TZ defines and uses a two-step handshake that realizes broker and client authentication with end-to-end encryption from the client to the TEE. The handshake protocol requires minimal pre-provisioned cryptographic material. The broker (server in TLS nomenclature) authentication is done through TLS' handshake, supported by default in *mosquitto*. The client authentication is done through MQTT. It publishes its symmetric key, encrypted with the broker's TEE public key, to a specific

write-only topic. This TEE key-pair is generated at device start-up time (secure boot) and derived from a Hardware Unique Key (HUK).

Layered Encryption & Access Control Mechanisms. Once the initial handshake is finished, MQT-TZ uses a two-layer encryption mechanism. First, the client-broker link is protected by TLS within MQTT. Second, MQTT's payload field is encrypted using the clients' symmetric key. Then, data is re-encrypted in the TEE (explained next) and sent to destination over MQTT-TLS. Doing so, we achieve end-to-end security relying on TRUSTZONE as a secure proxy.

Payload Re-encryption. The core secure functionality implemented in MQT-TZ is the payload re-encryption. We link MQTT with a Trusted Application (TA) running inside the TEE that transfers the encrypted data to the Secure World, retrieves the origin and destination keys from secure storage, and re-encrypts the information. Currently, topic subscription lists and MQTT metadata are stored in a dedicated database (MQTT DB) in the REE. We plan on shadowing these structures and keeping them in the TEE.

Lightweight Cache. MQT-TZ embeds a lightweight cache that keeps the most re-cent keys in the TA's heap memory, and evicts the least used to persistent secure storage.

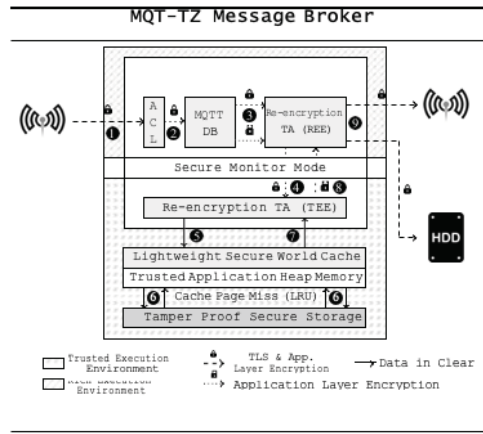
2.2 Implementation Details

MQT-TZ is implemented in C. The current version of MQT-TZ adds 400 SLOC to mosquitto and the TA amounts to 1184 SLOC. The MQT-TZ TA relies on OP-TEE (<https://www.optee.org>), an open-source framework with native support for TRUSTZONE. Our implementation will be publicly available (<https://github.com/mqtttz>).

Client and Server Authentication. The server-side authentication is done through vanilla TLS. We deploy MQT-TZ's secure broker in a device with a static IP address. Then, we bound the address to a domain name and use a certificate. We rely on *Let's Encrypt* (<https://letsencrypt.org/>) to get one and to authenticate the broker. The client-side authentication uses MQTT as communication layer, and openssl (v1.1.1a) for cryptographic primitives and operations. The integration with mosquitto exploits custom callbacks for each packet processing. In addition, we use MQTT Re-quest/Response (RR) features (since v5) for the client's key exchange. To control access and R/W permissions to topics, we use mosquitto's ACLs.

Trusted Application. We use OP-TEE to implement the payload re-encryption TA. Code developed within this framework has two parts: (1), a host app that runs in the REE and acts as entry point and bridge to the TEE, and (2) a trusted API in the TEE that exposes different functions. MQT-TZ intercepts all MQTT packets being forwarded to the recipients, and feeds our host app with both client's IDs, and the encrypted data. We then perform the

Figure 1. MQT-TZ Architecture and data flow.



payload re-encryption using OP-TEE's storage and cryptographic libraries. TRUSTZONE not only provides isolation between worlds, but also between different TAs. Hence, we use the same secure API to store new keys during the handshake. For the key retrieval, we plan to implement a small LRU cache to store the most frequently used keys in the TA's heap, and the rest in persistent secure storage.

3. Evaluation & Discussion

In this section, we perform an evaluation of MQT-TZ. First, we benchmark the TA re-encryption with random data in order to understand the overhead introduced by the re-encryption; and then, we analyze the CPU and network throughput when monitoring vital signs in a real setting. For all experiments, we virtualize a Raspberry Pi 3 using QEMU-V8 (<https://www.qemu.org/>) running mosquitto v1.6.3 and OP-TEE v3.5.0.

3.1 TA Re-encryption

In Fig. 2, we show the breakdown of the time required to re-encrypt a single block of data for different sizes. The time is split in the time to retrieve each key (retrieve dec key, retrieve enc key), encrypt, and decrypt. We can observe that AES is two orders of magnitude slower in the TEE. This is a consequence of OP-TEE not using hardware accelerators in contrast to openssl in the REE. Moreover, we observe sensible slowdowns when switching from in-memory to secure persistent storage.

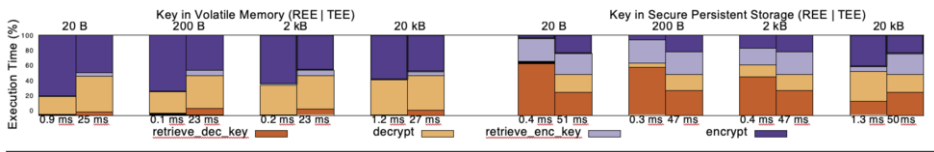


Figure 2. Re-encryption TA microbenchmark

3.2 Realtime ECG processing

In this case, we test the resilience of MQT-TZ at sustaining the workloads that can be encountered in a hospital. For the experiments, we use the LTMS-S [16] platform developed by CSEM for the European Space Agency (ESA). In particular, we simulate 50 patients streaming in real-time 1-lead electrocardiograms (ECGs) at a frequency of 321.25 Hz. All ECGs are streamed toward a single MQT-TZ broker. In Fig. 3, we depict the outbound throughput generated by each publisher measured using nethogs (<https://github.com/raboof/nethogs>). We observe that at any given time only a subset of the publishers actually emits data. A single subscriber streams at 350 Bytes/s in the worst case, and the full collective generates between 3 to 5 kBytes per second. During the experiment, we recorded using

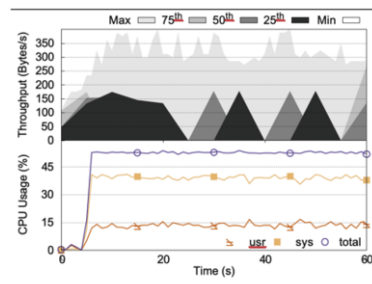


Figure 3. Workload test: Network throughput (top) and CPU usage (bottom)

dstat (<https://github.com/dagwieers/dstat>) the CPU load at the broker, shown in Fig. 3. We observe that after the initial peak, the overall CPU usage (both for usr and sys processes) stabilizes at 60%.

4. Conclusion and Future Work

Motivated by the lack of secure-by-design communication protocols for the Edge, we presented MQT-secure implementation of the MQTT broker using TRUST-ZONE and showed its direct application in a in-hospital setting. The proposed system is robust and capable of managing 50 patients in real-time with a CPU usage of 60%. We plan to extend this work along the following directions. First, we will compare MQT-TZ against other publish-subscribe protocols and messaging queues. Second, we will study the performance overhead of MQT-TZ when deployed on large-scale scenarios. Finally, we intend to look into the energy footprint, an aspect of paramount relevance for edge deployments.

References

- [1] G. P. Cumming, "Connecting & collaborating - Healthcare for the 21st century," in *PAHI'2014*, 2014.
- [2] J. Dunn, R. Runge, and M. Snyder, "Wearables and the medical revolution," *Pers. Med.*, vol. 15, no. 5, pp. 429–448, 2018.
- [3] M.-P. Gagnon, P. Ngangue, J. Payne-Gagnon, and M. Desmartis, "m-Health adoption by healthcare professionals: A systematic review," *J. Am. Med. Inform. Assn.*, vol. 23, pp. 212–220, June 2015.
- [4] The European Parliament and the Council of the European Union, "Regulation (EU) 2016/679," 2016.
- [5] A. Banks and R. Gupta, "MQTT version 3.1.1," software, OASIS, Oct. 2014.
- [6] Personal Connected Health Alliance, "Fundamentals of medical-grade data exchange," white paper, Continua, Sept. 2018.
- [7] IHE PCD Technical Committee, "Medical equipment management (MEM): Medical device cyber security," white paper, IHE International, Inc., Oct. 2015.
- [8] T. Yokotani and Y. Sasaki, "Comparison with HTTP and MQTT on required network resources for IoT," in *ICCEREC'2016*, pp. 1–6, Sept. 2016.
- [9] K. Chooruang and P. Mangkalakeeree, "Wireless heart rate monitoring system using MQTT," *Procedia Comput. Sci.*, vol. 86, pp. 160–163, 2016.
- [10] C. Segarra, R. Delgado-Gonzalo, M. Lemay, P.-L. Aublin, P. Pietzuch, and V. Schiavoni, "Using trusted execution environments for secure stream processing of medical data," in *Lect. Notes Comput. Sc.*, vol. 11534, pp. 91–107, 2019.
- [11] A. Krylovskiy, M. Jahn, and E. Patti, "Designing a smart city internet of things platform with microservice architecture," in *FiCloud'2015*, pp. 25–30, Aug. 2015.
- [12] Y. Lee, H. Hsiao, C. Huang, and S. T. Chou, "An integrated cloud-based smart home management system with community hierarchy," *IEEE. T Consum. Electr.*, vol. 62, pp. 1–9, Feb. 2016.
- [13] Teserakt AG, "Is MQTT secure? (a report)," 2019.
- [14] R. Liu and M. Srivastava, "VirtSense: Virtualize sensing through ARM TrustZone on Internet-of- Things," in *SysTEX'2018*, (New York, NY, USA), pp. 2–7, ACM, 2018.
- [15] J. Amacher and V. Schiavoni, "On the performance of ARM TrustZone," in *DAIS'2019*, pp. 133–151, 2019.
- [16] O. Chételat, D. Ferrario, M. Proença, J.-A. Porchet, A. Falhi, O. Grossenbacher, R. Delgado-Gonzalo, N. Della Ricca, and C. Sartori, "Clinical validation of LTMS-S: A wearable system for vital signs monitoring," in *EMBC'2015*, pp. 3125–3128, 2015.

Near Real Time EHR Data Utilization in a Clinical Study

Melody L. PENNING^{1a}, Collette BLACH^b, Anita WALDEN^a, Pei WANG^a, Katrina M. DONOVAN^a, Maryam Y. GARZA^a, Zhan WANG^a, Julie FRUND^a, Shorabuddin SYED^a, Mahanazuddin SYED^a, Guilherme DEL FIOLE^c, L. Kristin NEWBY^d, Carl PIEPER^e, Meredith ZOZUS^{a,f}

^aUniversity of Arkansas for Medical Sciences, Little Rock, AR, USA,

^bDuke University Molecular Physiology Institute, Durham NC, USA,

^cDepartment of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

^dDuke Clinical & Translational Science Institute, Duke University Medical Center, Durham, NC, USA

^eDepartment of Biostatistics & Bioinformatics, Duke University, Durham, NC, USA

^fUniversity of Texas Health Science Center San Antonio, San Antonio, TX, USA

Abstract. Extraction and use of Electronic Health Record (EHR) data is common in retrospective observational studies. However, electronic extraction and use of EHR data is rare during longitudinal prospective studies. One of the reasons is the amount of processing needed to assess data quality and assure consistency in meaning and format across multiple investigational sites. We report a case study of and lessons learned from acquisition and processing of EHR data in an ongoing basis during a clinical study.

Keywords. Data integration, clinical trials, data quality, data processing

1. Introduction

Large sets of electronic health record (EHR) data present an exciting possibility for research such as epidemiologic and observational research, real-time data collection, safety surveillance and regulatory uses and prospective clinical research. [3,4,6] Although this data has much potential, there are many issues that must be addressed in order to produce consistent and reliable findings. [10] EHR data are data collected in the process of care. Since the primary purpose of EHRs is to support activities such as clinical workflow and billing, the resulting data format, structure and quality often differ from the carefully scheduled, collected and formatted data required for research. In addition, EHR data changes over time reflecting (1) changes in the health status of the patient, and (2) changes in technology, data representation formats, regulatory incentives and institutional factors. [5] These variations in expected data versus collected data (e.g., timeliness, accuracy) are collectively referred to as data quality

¹ Corresponding Author, Melody L. Penning, University of Arkansas for Medical Sciences, Little Rock, AR, USA; E-mail: mlpenning@uams.edu.

dimensions. [8,11]

EHR data provide useful information about a patient and their healthcare only if the data quality is high and correspond to the real world. This correspondence varies widely, as do the measurement methods employed in the published literature. [1; 9] Further, the reasons for observed inconsistencies are rarely reported. To address this deficit in our understanding of data quality in healthcare data, the Patient Centered Outcomes Research Institute (PCORI) EHR Data Quality Study compares EHR and patient self-reported data for 34 conditions, 8 medical procedures, hospitalizations, and smoking status in cohorts from two southern U.S. states (Arkansas and North Carolina). This report outlines the data acquisition and processing of EHR data for this study. Briefly, the study required near real time acquisition and processing of EHR data to facilitate the comparison of EHR with self-report data to (1) identify discrepancies between these sources, and (2) support subsequent EHR data investigations and patient interviews on a weekly basis.

2. Methods

We report on the common aspects of data acquisition and processing between the two regions. Inpatient data from large hospitals were extracted from institutional data warehouses. Data from outpatient clinics were obtained either directly from the clinic EHR (North Carolina) or from an institutional data warehouse (Arkansas). Patient self-report data were compared with phenotypes derived from EHR data for the conditions, procedures, hospitalizations and smoking status from Arkansas and conditions from North Carolina.

For each phenotype, computational algorithms were developed to render a “self-report” answer and an “EHR answer”. Each was classified as “yes/confirmatory”, “don’t know/possible”, or “no/no evidence” in the self-report and EHR data respectively. Mismatches between EHR and self-reported phenotypes were considered a discrepancy. A subset of study participants were interviewed about their discrepancies. In the Arkansas cohort, participants were prospectively enrolled and the study required acquisition and processing of data on a weekly basis. In North Carolina, participants from a longitudinal community registry and biorepository, the Measurement to Understand Reclassification of Disease of Cabarrus/Kannapolis (MURDOCK) Study, have provided self-report data with annual updates as well as consent to access their EHR data. [2]

3. Results

For the North Carolina cohort, self-report data were managed in a custom-developed system. Existing self-report data were compared to newly acquired EHR data and discrepancy reports. For the Arkansas cohort, participants were prospectively enrolled at six clinics throughout the state. In Arkansas, self-report data were collected only once for the purpose of the PCORI EHR Data Quality Study. Data for the Arkansas cohort were managed in the open-source Community Edition of the OpenClinica Electronic Data Capture (OC-EDC) system. Discrepancies for both cohorts were double coded for root cause and correct data source in the OC-EDC system. The orchestration of these sources was accomplished with carefully timed data transfers and

SQL queries running over the entirety of the combined patient populations with 1,061,999 patients and 23,839,223 encounters.

Research Staff had real time access to EHR Data. For the Arkansas cohort, participant healthcare data were captured in either the Epic® EHR or various versions of Centricity®. These data were directly accessed through the EHR user interface by the research assistants (RAs) at enrollment to obtain participant medical record numbers (MRN) to for patient matching between self-report and EHR data. The RAs also used Epic to look up diagnostic information (1) prior to the telephone discrepancy interview when a discrepancy appeared likely due to clinical charting under the wrong patient or record linkage errors; and (2) for post interview attempts to confirm root cause. Because the study was operationalized differently in North Carolina due to previously existing EHR and self-report data, and because the clinics were separate institutions from the MURDOCK Study, EHRs in North Carolina were not accessed through the EHR user interface.

Acquisition of EHR data required substantial preprocessing. For the Arkansas cohort, all data were extracted from the institutional data warehouse for enrolled participants. The extracted data were stored in Microsoft SQL server and required several layers of conversion processing in order to integrate the regional program data with the hospital data. For the North Carolina cohort, EHR data were obtained from five local healthcare facilities with the following EHRs: Athena Health®, Allscripts®, Practice Fusion® and Epic®. The data from the different facilities were standardized and transformed to a common data model by the MURDOCK Study team. These data were linked with the self-report data, de-identified and sent to the study coordinating center for further processing. Record linkage between the incoming self-report data and EHR data was a deterministic match for both cohorts using the MURDOCK Study Number (North Carolina) and the MRN (Arkansas). Any linkage errors discovered in the EHR were collected as part of the study.

Following record linkage, each participant's EHR data were phenotyped. Phenotypes are rule-based definitions used for classification. Combinations of standardized controlled coded terminologies for healthcare (the International Classification of Diseases (ICD) and Current Procedural Terminology (CPT)), along with a few other non-coded data points identified from the existing literature and from practicing physicians, were used to develop the rule- based definitions. These were run over the entire dataset.

The phenotyping process was completed using stored procedures and each processing step was logged in case of failure. Overall, 140 phenotypes (70 "confirmatory" and 70 "possible") were specified and programmed. A total of 7,405 unique ICD-9 and ICD-10 codes were used in the process. Due to the long list of codes, a flexible design was necessary. Value sets for all phenotype definitions were kept in one table to be read by stored procedures, providing a quick and easy way to address unexpected phenotype results with simple table insertions/deletions as opposed to time consuming and risky code alterations. Records returned from the EHR for each phenotype run were compared with those from the self-report dataset. Each condition, procedure and event of interest for each participant was classified as "discrepant", "possibly discrepant" or "not discrepant". Discrepancies were then output to reports for RAs to use in the interview process. In addition to the discrepancy details, supplemental information such as ICD code descriptions were included to help characterize each discrepancy. Discrepancies were also imported into the OpenClinica system for post- interview coding (root cause and most likely correct data source) and

adjudication. Finally, coded and adjudicated discrepancies were exported from the OpenClinica system back to the Microsoft SQL Server for final post-interview discrepancy reports which were made available to participants.

To date, 2,151 participants have been enrolled in Arkansas. Of those, 1,096 participants have been successfully processed through discrepancy identification. Of the North Carolina cohort, 3,927 MURDOCK Study participants had EHR data and have been processed through discrepancy identification, of which 373 participants have been interviewed. The study is ongoing at this time.

4. Discussion

The lack of information about the level of correspondence between healthcare data and the real world is an impediment to various types of research. Data quality must not be an unknown. The PCORI EHR Data Quality Study is an effort to test dimensions of healthcare data quality over multiple phenotypes simultaneously in order to reveal data quality issues at scale and establish general measures and methods for characterization of data quality problems in health record data.

Modifications were needed in response to both internal and external changes. Some phenotypes were especially challenging, such as obesity, kidney disease and chronic obstructive pulmonary disease (COPD). Initially phenotypes were developed with broad definitions and code sets. Through iterations of ICD code tuning they were refined to more conservative sets that reflected the intent of the phenotype. In addition, an inherent dependency on data source systems provided challenges. Changes to institutional software along with a national migration from ICD version 9 to version 10 affected our design and led to reanalysis, redefinition and reprogramming. During system development, a new EHR system for the Arkansas regional clinics was implemented which required additional programming to access the data that had not yet been integrated into the institutional data warehouse. Additionally, some specific ICD codes were also updated during this time period. System changes are inevitable, as institutions maintain EHR systems and controlled terminologies evolve. Due to flexibility of the system design, we were able to resolve phenotype tuning issues and adapt to external and internal environmental changes smoothly.

Several concrete lessons were learned that are unique to comparing EHR with self-report data. First, record linkage can be problematic, even in cases where id numbers are used. [7; 8] Linked data should always be fact checked to make sure errors have not occurred. Second, phenotype definitions will evolve and change based on data source, date of collection and workflow or process of collection. Third, comparisons for events will cause a multiplicity of “no match” conditions (e.g. hospitalizations and procedures can easily lead to a large *all possible combinations* problem if the date windows are not sufficiently tolerant). Lastly, the permutations of possible outcomes (discrepant, possible, and not discrepant) must be very carefully considered. These conditions were discussed on multiple occasions as new information came in from phenotype runs and were compared directly against data in Epic®.

This work has provided a window into the quality of EHR data through the design and implementation of a near-real-time data quality measurement system while supporting study operations for informatics and research. While the architecture was designed for the University of Arkansas’ environment it may be adapted and generalized to work at other institutions and for other data sources.

5. Conclusion

It is well-known that the quality of EHR data varies from institution to institution. The implementation reported here identified additional challenges in acquisition and assessment of discrepancies between EHR data and patient self-report data. Even with well-defined approaches a significant amount of effort to adapt and keep up with institutional changes and handle unpredictable quality issues was needed. The work reported here required a team- based approach with the technical staff, informaticists, clinicians and other subject matter experts. Observations and lessons learned from this implementation may inform others as they embark upon use of EHR data during and to support clinical and informatics studies.

6. Funding Acknowledgement.

Funding was provide by PCORI (ME-1409-22573), the David H. Murdock Institute and the National Institutes of Health, National Library of Medicine (K99LM011128, R00LM011128). The ideas and opinions expressed here are those of the authors and not necessarily those of PCORI, the National Institutes of Health or any other organization.

References

- [1] H. Baumeister, et al., High agreement of self-report and physician-diagnosed somatic conditions yields limited bias in examining mental–physical comorbidity, *Journal of clinical epidemiology* **63** (2010), 558-565.
- [2] S. Bhattacharya, et al., The Measurement to Understand Reclassification of Disease of Cabarrus/Kannapolis (MURDOCK) study community registry and biorepository, *American Journal of Translational Research* **4** (2012), 458.
- [3] M.R. Cowie, et al., Electronic health records to facilitate clinical research, *Clinical Research in Cardiology* **106** (2017), 1-9.
- [4] R.L. Fleurence, et al., Launching PCORnet, a national patient-centered clinical research network, *Journal of the American Medical Informatics Association* **21** (2014), 578-582.
- [5] G. Hripcsak, D.J. Albers, High-fidelity phenotyping: richness and freedom from bias, *J Am Med Inform Assoc* (2017).
- [6] J.M. Overhage, L.M. Overhage, Sensible use of observational clinical data, *Statistical methods in medical research* **22** (2013), 7-13.
- [7] M. Penning, Inferred Error Rates for Entity Resolution in Healthcare, in: *The International Conference on Information Quality (MIT ICIQ)*, Little Rock, 2017.
- [8] J.R. Talburt, et al., Evaluating and Improving Data Fusion Accuracy, in: *Information Quality in Information Fusion and Decision Making*, É. Bossé and G.L. Rogova, eds., Springer International Publishing, Cham, 2019, pp. 295-326.
- [9] K. Wada, et al., Self-reported medical history was generally accurate among Japanese workplace population, *Journal of clinical epidemiology* **62** (2009), 306-313.
- [10] N.G. Weiskopf, C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, *Journal of the American Medical Informatics Association* **20** (2013), 144-151.
- [11] M.N. Zozus, et al., Factors impacting physician use of information charted by others, *JAMIA Open* (2018).

Negation Detection for Clinical Text Mining in Russian

Anastasia FUNKNER^{a, 1}, Ksenia BALABAEVA^a, and Sergey KOVALCHUK^a
^a*ITMO University, Saint Petersburg, Russia*

Abstract. Developing predictive modeling in medicine requires additional features from unstructured clinical texts. In Russia, there are no instruments for natural language processing to cope with problems of medical records. This paper is devoted to a module of negation detection. The corpus-free machine learning method is based on gradient boosting classifier is used to detect whether a disease is denied, not mentioned or presented in the text. The detector classifies negations for five diseases and shows average F-score from 0.81 to 0.93. The benefits of negation detection have been demonstrated by predicting the presence of surgery for patients with the acute coronary syndrome.

Keywords. clinical texts, medical records, electronic medical records, natural language processing, negation detection, Russian, anamnesis

1. Introduction

An electronic medical record (EMR) contains records in natural language, for example, life and disease anamnesis, protocols of surgeries, examination or discharge reports. Developing predictive modeling in medicine requires additional features, and not only strictly structured ones, such as blood tests, patient metrics, etc. For the English language, there are many tools for labeling text, extracting entities, disease's cases, temporal and negation detection (UMLS, UIMA, IBM Watson, Apache Ruta, etc.) [1–4]. However, most of the tools are now impossible to adapt to the Russian language because a word corpus of medical texts has not yet been compiled. There is a small one based on 120 records with labeled diseases and their attributes (complications, severity, treatment, etc.) [5,6]. But this corpus is compiled on the records of one medical centre and for patients with a limited set of diseases, and therefore it is difficult to use it for automatic text labeling of other medical centres. Thus, our team aims to develop a system of four modules to solve problems of misprints in medical terms, negation, temporality, and experimenter detection (see Figure 1). This paper is devoted to a module of negation detection.

It is accepted that any condition or disease is sought by the occurrence of synonyms-terms in medical history, anamnesis, etc. However, the presence of such a term may also indicate that the patient denies the disease. In Russia, healthcare providers often ask patients after admission to the hospital about strokes, heart attacks,

¹ Corresponding Author, ITMO University, 49 Kronverksky Pr., St. Petersburg, 197101, Russia; E-mail: funkner.anastasia@itmo.ru, funkner.anastasia@gmail.com.

and high blood pressure in the past, as this may affect the course of treatment. As a result, in disease anamnesis, phrases like “MI in the past denies” (Myocardial Infarction), “there was no increase in blood pressure”, “without a history of hypertension”, “diabetes denies” can be often found.

The above-mentioned problem is called negation detection in texts. One of the first algorithms NegEx was proposed by Chapman et al. in 2001 and was based on rules [7]. Subsequently, syntax-based methods for the English language were developed [8,9]. Goryachev et al. implemented four methods for detecting negations and compared them using three sets of clinical records in English [10]. NegEx showed the best accuracy [7]. Thus, most of the approaches found for identifying negations are developed with grammatical rules for the English language.

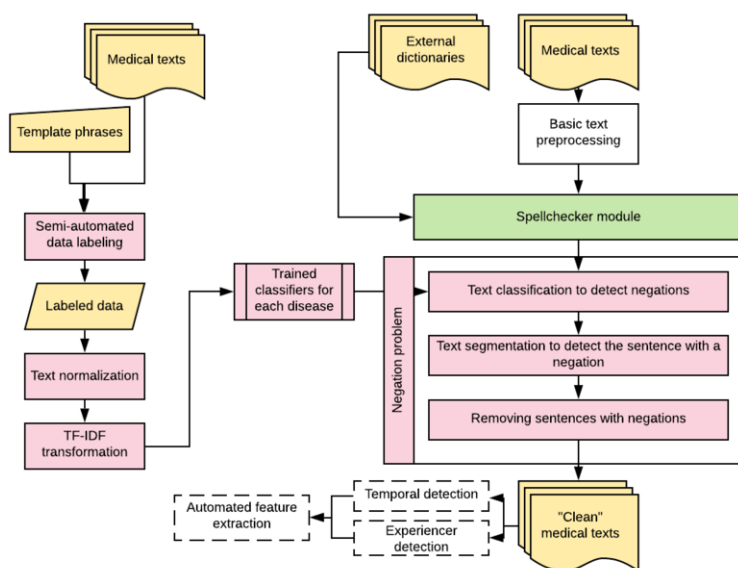


Figure 1. Modules of spelling correction and negation detection. Pink blocks indicate methods which are described in this paper. Green blocks are developed and implemented by colleagues. Dashed blocks indicate methods that are just being developed.

2. Method

In this research, we used a set of anonymized 3434 EMRs of patients with the acute coronary syndrome (ACS) who admitted to Almazov National Medical Research Centre (Almazov Centre) during 2010-2015. Disease anamneses are one of the most unstructured records (free text without any tags; each physician writes as he/she wants), and therefore we used them to demonstrate our approach.

Our approach to detect negations is machine learning based on multi-class classification. In Figure 1 the main steps are presented on how to train a negation detector and use for new data sets. Firstly, clinical texts should be annotated with 3 labels for each specific disease: ‘-1’ means ‘the disease is denied’, ‘0’ means ‘the disease is not mentioned in the text’ and ‘1’ means ‘patient has this disease’. We suppose to annotate the whole anamnesis with these labels because the classifier will be

able to learn an additional context which follows the disease. The two problems are possible: the disease can relate to a patient’s family members (label with 0) and there can be two descriptions about patient’s condition in different moments of her or his life (label with 1). We use semi-automated labeling with phrase search because the provided anamneses can be written in a similar format by the same doctor and it is possible to find repeated phrases.

Secondly, annotated anamnesis is normalized with pymorphy2 module which can be used for Russian and Ukraine [11]. Term frequency – Inverse document frequency (TF-IDF) transformation makes commonly used words less weighted in word corpus (prepositions, pronounce, etc.). TF-IDF vectors are used to train, tune and chose the best classifier. In consequence of cross-validation experiments, based on decision trees a gradient boosting classifier shows the best result with F-score. Thus, we trained five classifiers for detecting arterial hypertension (AH), myocardial infarction (MI), stroke, diabetes mellitus (DM) and angina pectoris (AP) (see Table 1). Using negation classifiers is crucial for stroke, MI and AH. For MI and AH, the classifiers learn context and help find additional cases of these conditions. Table 2 includes the most import phrases for detecting MI which are disease terms and possible treatment (surgery and medications).

When the negations are detected, each sentence of the anamnesis is labeled as containing or not containing negation using a logistic loss [12]. Sentences or their parts with negations are removed from anamnesis so that these texts can be used to build other models and consider only the patient's existing conditions, for instance, for topic modeling.

Table 1. The comparative results for negation classifiers (F1 – F-score). False positives are anamneses which were detected with ‘1’ and ‘-1’ using a term search and a negation classifier respectively. False negatives were detected with ‘0’ (terms not found in a text) and ‘1’ with the same methods.

Disease	Number of annotated anamneses	F1 for ‘-1’ labeled	F1 for ‘0’ labeled	F1 for ‘1’ labeled	Macro average F1	Total accuracy	Number of false positives (%)	Number of false negatives (%)
stroke	648	0.96	0.96	0.78	0.9	0.93	0.15	0.00
MI	897	0.95	0.5	0.97	0.81	0.96	0.07	0.15
AH	1167	0.88	0.59	0.96	0.81	0.93	0.08	0.20
DM	239	0.95	0.93	0.91	0.93	0.93	0.02	0.00
AP	864	0.67	0.91	0.98	0.85	0.95	0.02	0.01

Table 2. Feature importance (FI) of words and phrases which are used to define the presence of disease in an anamnesis.

Russian phrase	English phrase	FI	Class	Comments
оим	AMI	0.052	1	Abbreviation: Acute Myocardial Infarction
ибс оим	IHD AMI	0.042	1	Abbreviation: Ischemic Heart Disease. IHD includes myocardial infarction.
миокард отрицать	myocardial deny	0.041	-1	Patient denies MI
отрицать	deny	0.035	-1	Patient denies an acute condition
миокард ранее	myocardial early	0.027	1	Indicates myocardial problems previously
нижний	low	0.022	1	Indicates the localization of AMI
провести стентирование	perform stenting	0.016	1	Stenting is a surgery performed after AMI.
ознакомить	provide	0.015	0	Indicates anamnesis without a medical history

передний	anterior	0.014	1	Indicates the localization of AMI Medicines that are used to treat AMI
анальгин	analgin, tramadol	0.013	1	
трамадол				

3. Results

To evaluate how the application of negation detection module affects predictive model performance we conduct experiments on 2 tasks: clinical episode outcome prediction and prediction of surgery for the patients suffering from ACS [13]. Usually, all information medics have about a patient by the time of arrival is clinical anamnesis in the natural text format. Therefore, we use it as input data to train machine learning models.

During experiments all the models parameters remain constant (for XGBoost: num. estimators = 500, max depth = 100, learning rate = 0.1, objective = binary logistic; for Random Forest: max depth = 100, num estimators = 500, criterion = 'gini', min. samples split = 2, min. samples leaf = 1; for k-nearest neighbors model (KNN): num. neighbours = 10, weights = 'uniform'). The only change is in the input data for the fixed feature set: we use diagnosis features collected from clinical texts using regular expressions without negation detector and with the help of negation detector.

We evaluate the results on 33% test sample using the F1 score as a quality metric. The results are performed in Table 3.

Table 3. Evaluation of the negation detection (ND) performance on predictive tasks.

Task	Model	F1 without ND Text features	F1 with ND Text features	F1 without ND Text + other features	F1 with ND Text + other features
Surgery prediction	XGBoost	0.3755	0.5234	0.6941	0.6992
	Random Forest	0.3948	0.7110	0.7038	0.7128
	KNN	0.5091	0.5234	0.5851	0.5870
Outcome prediction	XGBoost	0.3234	0.3233	0.5208	0.5113
	Random Forest	0.3234	0.4229	0.4402	0.4569
	KNN	0.3234	0.3233	0.3428	0.3428

4. Discussion

According to the experiment results, we may conclude that the use of negation detector significantly improves the performances of XGBoost, Random Forest and KNN on the task of surgery prediction based on only text features. On the task of the outcome prediction, only Random forest improves. All other algorithms' results do not change. One of the reasons could be the need for additional information for the outcome prediction to improve since on current feature sample models predict the constant class for all instances and can't properly distinguish classes. For this reason, we conduct additional experiments and add other numerical features based on the first lab test the patients had. Lab tests for models are more significant than texts, therefore, the effect from the negation detection application is not such significant in this experimental set.

5. Conclusion

We demonstrated the procedure for training and using the negation detector on the example of disease anamnesis of patients with ACS. For five treatment-relevant diseases and conditions, the negation detector showed average F-score from 0.81 to 0.93 on the test sample. Also, the benefits of the negation detection were demonstrated with predicting the presence of surgery for patients with ACS. The developed and implemented module for detecting and removing negations is part of the application for processing clinical text. In the future, we plan to develop additional modules and make this application an indispensable part of data preparation and feature extraction for any predictive modeling.

Acknowledgements

This research is financially supported by The Russian Science Foundation, Agreement #19-11-00326.

References

- [1] J. Jonnagaddala, S.T. Liaw, P. Ray, M. Kumar, N.W. Chang, and H.J. Dai, Coronary artery disease risk assessment from unstructured electronic health records using text mining, *J. Biomed. Inform.* 58 (2015) S203–S210. doi:10.1016/j.jbi.2015.08.003.
- [2] N. Mehta, and M. V. Devarakonda, Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey?, *J. Allergy Clin. Immunol.* 141 (2018) 2019–2021.e1. doi:10.1016/j.jaci.2018.02.025.
- [3] E. Ford, J.A. Carroll, H.E. Smith, D. Scott, and J.A. Cassell, Extracting information from the text of electronic medical records to improve case detection: A systematic review, *J. Am. Med. Informatics Assoc.* 23 (2016) 1007–1015. doi:10.1093/jamia/ocv180.
- [4] C.C. Huang, and Z. Lu, Community challenges in biomedical text mining over 10 years: Success, failure and the future, *Brief. Bioinform.* 17 (2016) 132–144. doi:10.1093/bib/bbv024.
- [5] A.O. Shelmanov, I. V. Smirnov, and E.A. Vishneva, Information extraction from clinical texts in Russian, *Komp'yuternaja Lingvistika i Intellektual'nye Tehnol.* 1 (2015) 560–572.
- [6] B. A.A., N.-B. L.S., N. I.V., D. D.A., A.O. Shelmanov, and V. E.A., Technologies for Complex Intelligent Clinical Data Analysis, *Вестник Российской Академии Медицинских Наук.* 71 (2016) 160–171. doi:https://doi.org/10.15690/vramn663.
- [7] W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, and B.G. Buchanan, A simple algorithm for identifying negated findings and diseases in discharge summaries, *J. Biomed. Inform.* (2001). doi:10.1006/jbin.2001.1029.
- [8] Y. Huang, and H.J. Lowe, A Novel Hybrid Approach to Automated Negation Detection in Clinical Radiology Reports, *J. Am. Med. Informatics Assoc.* 14 (2007) 304–311. doi:10.1197/jamia.M2284.
- [9] S. Mehrabi, A. Krishnan, S. Sohn, A.M. Roch, H. Schmidt, J. Kesterson, C. Beesley, P. Dexter, C. Max Schmidt, H. Liu, and M. Palakal, DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx, *J. Biomed. Inform.* 54 (2015) 213–219. doi:10.1016/j.jbi.2015.02.010.
- [10] S. Goryachev, M. Sordo, Q. Zeng, and L. Ngo, Implementation and Evaluation of Four Different Methods of Negation Detection, Boston, MA DSG. (2006) 2826–2831. doi:10.2307/2283959.
- [11] M. Korobov, Morphological analyzer and generator for Russian and Ukrainian languages, in: *Commun. Comput. Inf. Sci.*, 2015. doi:10.1007/978-3-319-26123-2_31.
- [12] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, Explainable AI for Trees: From Local Explanations to Global Understanding, (2019) 1–72. <http://arxiv.org/abs/1905.04610>.
- [13] A. Funkner, S. Kovalchuk, and K. Bochenina, Preoperational Time Prediction for Percutaneous Coronary Intervention Using Machine Learning Techniques, in: *Procedia Comput. Sci.*, 2016: pp. 172–176. doi:10.1016/j.procs.2016.11.021.

On the Construction of Multilingual Corpora for Clinical Text Mining

Fabián VILLENA^{a,b}, Urs EISENMANN^a, Petra KNAUP^a, Jocelyn DUNSTAN^{b,c},
and Matthias GANZINGER^{a,1}

^a*Institute of Medical Biometry and Informatics, Heidelberg University, Germany*

^b*Center of Medical Informatics and Telemedicine, University of Chile, Chile*

^c*Center for Mathematical Modeling, University of Chile, Chile*

Abstract. The amount of digital data derived from healthcare processes have increased tremendously in the last years. This applies especially to unstructured data, which are often hard to analyze due to the lack of available tools to process and extract information. Natural language processing is often used in medicine, but the majority of tools used by researchers are developed primarily for the English language. For developing and testing natural language processing methods, it is important to have a suitable corpus, specific to the medical domain that covers the intended target language. To improve the potential of natural language processing research, we developed tools to derive language specific medical corpora from publicly available text sources. In order to extract medicine-specific unstructured text data, openly available publications from biomedical journals were used in a four-step process: (1) medical journal databases were scraped to download the articles, (2) the articles were parsed and consolidated into a single repository, (3) the content of the repository was described, and (4) the text data and the codes were released. In total, 93 969 articles were retrieved, with a word count of 83 868 501 in three different languages (German, English, and Spanish) from two medical journal databases. Our results show that unstructured text data extraction from openly available medical journal databases for the construction of unified corpora of medical text data can be achieved through web scraping techniques.

Keywords. Natural Language Processing, Data Mining, Information Storage and Retrieval, Medical Informatics, Linguistics

1. Introduction

The amount of digital data derived from healthcare processes have increased tremendously in the last years [11]. Data available from healthcare sources can be roughly divided into two types: (1) structured data, namely information already organized into a known data model, for instance, results from psychological surveys or results of laboratory tests; and (2) unstructured data in the form of free-text data, medical images, or physiological signals. Examples for such free-texts are anamnesis records or discharge

¹Corresponding Author: Institute of Medical Biometry and Informatics, Heidelberg University, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany; E-mail: matthias.ganzinger@med.uni-heidelberg.de.

notes and they are commonly written by healthcare professionals [7]. Dalianis et al. estimated that over 40% of healthcare data are unstructured [3]. It is often hard to use such unstructured texts for research and other analyses due to the lack of available tools to process and extract information. These tools must be tailored specifically for the medical field, and moreover, tailored for a specific language [3].

Nowadays, the way to analyze texts is using large corpora from which computers can learn the meaning of a word from its context, replacing former rule-based and feature-selection-based approaches [5]. However, availability of medical records for such research purposes is usually restricted due to privacy issues. Consequently, collecting biomedical texts extracted from journals have been proposed as a possible solution for the lack of resources [1, 12].

Natural language processing (NLP) is a sub-field of computer science, concerned with interactions between computers and human languages [4]. Some of the most important NLP applications in medicine are (1) detection of patterns in electronic health records (EHR) (e.g. detection of healthcare associated infections, adverse drug events or cancer symptoms); (2) text summarizing and translation of EHR; and (3) automatic codification [2]. While there are successful applications of NLP in the medical field [6, 13, 9], the majority of tools used by researchers are developed for the English language [8]. Therefore, creating tools and training corpora for languages different to English are valuable contributions to empower NLP of clinical sources in non-English speaking countries. To improve the potential of NLP research, especially for non-English languages, we developed tools to derive language specific medical corpora from publicly available text sources. Specifically, we focus on Spanish and German texts.

2. Methods

In order to extract medicine-specific unstructured text data, openly available papers from biomedical journals were used. The general process was divided into (1) web scraping of the database to download each of the articles contained in the website, (2) parsing of the raw downloaded data to consolidate it into a local standardized repository, (3) descriptive analysis of the content, and (4) the release of the data to the scientific community. A general overview of the workflow is summarized in Figure 1. The entire pipeline was developed in the Python² programming language.

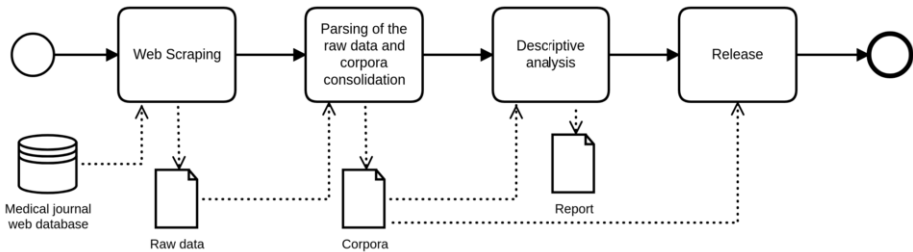


Figure 1. Overview of our proposed workflow to generate corpora from medical journal databases.

²Python - <https://www.python.org/>

2.1. Web scraping

Web scraping is a technique to extract automatically data from websites. The package Scrapy³ was used to develop a script, which navigates through the database and downloads every article in XML format.

The recursive algorithm first searched for a medical journal in the index of the database, then inside the journal looked up for the volumes available, and finally the script searched for all the articles inside that volume and downloaded the XML file associated with the article.

2.2. Parsing of the raw data and corpora consolidation

For parsing raw XML files, the package BeautifulSoup⁴ was used. Additionally, we extracted the following attributes: title, language, abstract, day of release and journal / meeting name.

The parsed articles were then consolidated into separate corpora based on the language in which the article was written. Each corpus was word tokenized using the NLP framework Spacy⁵.

2.3. Descriptive analysis

For the exploration of the corpora, each corpus was summarized by (1) the number of articles that is composed of, (2) number of word tokens, and (3) vocabulary size counted as the number of different words. To compare each of the corpora, we first deleted stop-words from the text using the list provided in Spacy and then frequency of appearance of each word in each corpus was calculated.

2.4. Release

The corpora are made available to the research community as a package, which contains each article labeled with the language of the content. The code used to retrieve the corpora is available on a git repository⁶ and the corpora in a hosting platform⁷.

3. Results

In total, 95 737 articles were retrieved from the German Medical Science Database⁸ (GMS) and the Chilean Scientific Electronic Library Online⁹ (SciELO). After parsing the raw data, 1 762 empty articles were discarded. A division by language was made, finding that 63% of the articles were written in German, 24% in English, 13% in Spanish and <1% in French. The metrics of each corpus are described in Table 1.

A word frequency analysis was made for each corpus and ten of the most frequent words are summarized in Table 2.

³Scrapy: A Fast and Powerful Scraping and Web Crawling Framework - <https://scrapy.org>

⁴Beautiful Soup - <https://www.crummy.com/software/BeautifulSoup>

⁵Spacy: Industrial-Strength Natural Language Processing - <https://spacy.io>

⁶<https://github.com/fvillena/multilingual-medical-nlp>

⁷<https://zenodo.org/record/3463379.XY4RsUEzaV4>

⁸<https://www.egms.de>

⁹<https://scielo.conicyt.cl>

Table 1. Summary statistics of the corpora

Metric	corpus		
	German	English	Spanish
Articles count	59 539	22 372	12 058
Number of word tokens	20 437 502	12 093 145	51 337 854
Vocabulary size	497 256	144 550	374 877

Table 2. Most frequent words in the corpora

Rank	corpus		
	German	English	Spanish
1	patienten (<i>patients</i>)	patients	pacientes (<i>patients</i>)
2	ergebnisse (<i>results</i>)	results	estudio (<i>study</i>)
2	methoden (<i>methods</i>)	study	años (<i>years</i>)
3	schlussfolgerung (<i>conclusion</i>)	treatment	casos (<i>cases</i>)
4	einleitung (<i>introduction</i>)	methods	tratamiento (<i>treatment</i>)
5	fragestellung (<i>research question</i>)	clinical	diagnóstico (<i>diagnostic</i>)
6	studie (<i>study</i>)	medical	riesgo (<i>risk</i>)
7	material (<i>material</i>)	group	figura (<i>figure</i>)
8	hintergrund (<i>background</i>)	used	enfermedad (<i>disease</i>)
9	therapie (<i>therapy</i>)	years	resultados (<i>results</i>)
10	durchgeführt (<i>carried out</i>)	time	forma (<i>form</i>)

4. Discussion

Current research on NLP in medicine mainly focuses in the English language; nevertheless there is an increasing effort in closing the gap of the development of NLP in languages beyond English [8]. Our work is a building block of a framework to retrieve medical unstructured text by the construction of biomedical corpora to support clinical text mining. Similar work was published by Soares et al. [10], but only SciELO database was scraped.

The toolset we describe in this manuscript has shown to be suitable for establishing generic medical corpora for multiple languages. To validate our approach we have queried a public Chilean and German database for medicine related publications. While we were focusing on German and Spanish texts, we retrieved a considerable amount of English texts because many the German site publishes in English as well.

Analysis has shown that the most frequent words conform across the corpora only partially (cf. Table 2). This result is to some extent unexpected, since scientific texts, as they are provided by the the underlying collections, typically share a very similar structure (IMRaD – Introduction, Methods, Results, and Discussion). While the reason for this finding needs further investigation, it might be due to differing focus areas in the collections. This result is supported by the figures shown in Table 1. Obviously, the GMS database contains shorter texts, possibly due to a large amount of conference abstracts published by them.

With a large amount of medical text data, the simplest model we can develop is a probability distribution, where we can predict upcoming words given a sentence. These models are essential in tasks where we have noisy and ambiguous inputs [4], as they oc-

cur, for instance, in spelling and syntax correction of clinical texts or speech recognition tasks for electronic health records (EHR).

5. Conclusion

The unstructured text data extraction from openly available medical journal databases for the construction of unified corpora of medical text data can be achieved through web scraping techniques.

References

- [1] Zhiwei Chen et al. "Evaluating semantic relations in neural word embeddings with biomedical and general domain knowledge bases". en. In: *BMC Medical Informatics and Decision Making* 18.S2 (July 2018). (Visited on 07/19/2019).
- [2] Hercules Dalianis. *Clinical Text Mining*. en. Cham: Springer International Publishing, 2018. (Visited on 09/23/2019).
- [3] Hercules Dalianis, Martin Hassel, and Sumithra Velupillai. "The Stockholm EPR Corpus – Characteristics and Some Initial Findings". en. In: *Management Research* (2009), p. 7.
- [4] Dan Jurafsky and James H. Martin. *Speech and Language Processing*. 3rd ed. draft. 2018.
- [5] Yoav Goldberg. "A Primer on Neural Network Models for Natural Language Processing". In: *CoRR* abs/1510.00726 (2015).
- [6] Kory Kreimeyer et al. "Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review". eng. In: *Journal of Biomedical Informatics* 73 (2017), pp. 14–29.
- [7] Clement J. McDonald. "The Barriers to Electronic Medical Record Systems and How to Overcome Them". In: *Journal of the American Medical Informatics Association* 4.3 (May 1997), pp. 213–221.
- [8] Aurélie Névél et al. "Clinical Natural Language Processing in languages other than English: opportunities and challenges". eng. In: *Journal of Biomedical Semantics* 9.1 (2018), p. 12.
- [9] Ewoud Pons et al. "Natural Language Processing in Radiology: A Systematic Review". eng. In: *Radiology* 279.2 (May 2016), pp. 329–343.
- [10] Felipe Soares et al. "Medical Word Embeddings for Spanish: Development and Evaluation". In: *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Minneapolis, Minnesota, USA: Association for Computational Linguistics, June 2019, pp. 124–133.
- [11] S. S. -L. Tan, G. Gao, and S. Koch. "Big Data and Analytics in Healthcare". en. In: *Methods of Information in Medicine* 54.06 (2015), pp. 546–547. (Visited on 09/23/2019).
- [12] Yanshan Wang et al. "A comparison of word embeddings for the biomedical natural language processing". In: *Journal of Biomedical Informatics* 87 (2018), pp. 12–20.
- [13] Wen-Wai Yim et al. "Natural Language Processing in Oncology: A Review". eng. In: *JAMA oncology* 2.6 (June 2016), pp. 797–804.

Ontology-Based and Architecture-Based Method for the Development of Interoperable Care Systems for Type 2 Diabetes Mellitus

Gustavo A. URIBE^{a,b,c,1}, Bernd BLOBEL^{d,e,f}, Diego M. LOPEZ^c,
Stefan SCHULZ^g, and Alonso A. RUIZ^h

^a*The European Organization for Nuclear Research, Geneva, Switzerland*

^b*Experimental High Energy Physics, University Antonio Nariño, Bogotá, Colombia*

^c*Telematics Engineering Research Group, University of Cauca, Popayán, Colombia*

^d*Medical Faculty, University of Regensburg, Regensburg, Germany*

^e*Health Competence Center Bavaria, Regensburg, Germany*

^f*First Medical Faculty, Charles University Prague, Prague, Czech Republic*

^g*Institute of Medical Informatics, Statistics and Documentation, Medical University of
Graz, Graz, Austria*

^h*Medical Faculty, University of Cauca, Popayán, Colombia*

Abstract. Processes like the care of type 2 diabetes mellitus patients require support by information systems considering the heterogeneity of the actors from different domains involved, enabling harmonization and integration of their specific methodologies and knowledge representation approaches towards interdisciplinary cooperation. Currently, the development of systems starts from the simplified information world, ignoring the aforementioned heterogeneity and specificity of real-world processes. This paper aims to demonstrate the feasibility of developing an adaptive, interoperable and intelligent system that supports the major aspects of type 2 diabetes mellitus care based on the Generic Component Model as formal methodology for modelling universal systems. The result is a deployable solution based on a formal representation of the diabetes care system, its objectives, and the intended business process. The implemented system enables reasoning over the data, inferring medical diagnosis. The effectiveness of the inference was evaluated, obtaining an F-measure of 0.89. The methods presented in this paper helps to build high quality models based on computation-independent aspects, which enable the construction of knowledge-based adaptive, intelligent and interoperable eHealth systems.

Keywords. Diabetes mellitus, Health information management, Ontology-based Systems, Systems architecture, Knowledge-based systems

1. Introduction

Currently, numerous software applications have been developed and implemented to support health care services. Few of them aim at achieving data interoperability, mapping

¹ Gustavo A. Uribe, ATLAS experiment, CERN, Espl. des Particules 1, 1211 Meyrin, Switzerland; E-mail: gustavo.uribe@cern.ch.

the structure of their data to different standards. Even fewer systems consider the concepts and contexts of the data represented, i.e., the knowledge of the actors and the interoperability problems associated with the heterogeneity of this knowledge. Therefore, most of the systems have difficulties to “understand” the data and to adapt them consistently and correctly to the different stakeholders. Improving health care services by interpreting data to perform conclusions and actions requires a specific level of system intelligence, especially needed in low and middle income countries where skilled human resources are limited.

The deployed methodology, i.e. the architectural model and framework, for designing and implementing a type 2 diabetes mellitus (T2DM) software system is based on the Generic Component Model (GCM) [1]. GCM allows the representation of any system architecture and the analysis and design of information systems. The GCM has been explained, discussed, used and compare in several papers, e.g. [1-4]. However, most of those papers focus on the principles GCM, without evaluating concrete software implementations. The proposed method allows the harmonization of the diversity of terms and concepts involved in health services by abstracting them. Thus, interoperability, collaboration and decision-making between health systems actors are demonstrated for a T2DM care scenario in the Colombian context. The paper’s main objective is to demonstrate the feasibility of developing an adaptive, interoperable and intelligent T2DM care system based on a formal methodology for modelling of systems and their ontological representation. The T2DM scenario was considered because of its high prevalence, 8.5% in 2014 [5], and its complex management due the required collaboration of many health professionals from different domains.

2. Methods

The methodology followed in this work is based on the GCM framework for analysis, design and implementation of the systems, providing a generic abstract architecture for each domain-specific aspect of the system instantiated by domain-specific ontologies. The obtained model addresses three domains, viz. the medical, policy and resource domain. The upper-level ontology BioTopLite [6] was deployed to harmonize the different domains’ concepts and knowledge. It provides high compatibility with the top-level ontology BFO [7], and it additionally considers some relevant and general aspects of the biological domain.

Ontology languages such as the Web Ontology Language (OWL) were used to represent architectural and static aspects of the system such as the generic system description and the description of the glycaemic control use case. Functional or dynamic aspects of the system are modelled by the Business Process Modeling Notation (BPMN) [8]. The business process model describes the expected behaviour of a business system and allows controlling, monitoring or implementing the ongoing processes.

Policies and rules governing the system for the glycaemic control use case were formally described using the SPARQL Inference Notation (SPIN) [9]. SPIN is a general purpose, standardized, and broadly accepted rule language that allows the definition of rules related to ontology classes. Now, the computation-independent T2DM care system was completely modelled. Based on this description, the software development process was performed according to the ISO 10746 RM-ODP [10] and its viewpoints as defined in the GCM. The output of this process is the specification of the information technology (IT) solution and, finally, the running application system.

3. Results

Generic and specialized architectural models of the system have been explained in [11-13]. For each GCM viewpoint, the models are adapted according to some inputs required for the development process [1, 13]. The Business View corresponds to the architectural models expressed using OWL and SPIN, and semi-formal models of the behaviour represented using BPMN. The Enterprise View defines the roles, activities and policies statements of the specified system [10]. The actors' roles in the system can be classified into: the categories of health organization staff, the self-care actor, the organizational administrator, and the resource chief. The roles have been derived from the International Standard Classification of Occupations [14]. The Information View represents in a platform-independent way the semantics of information [10]. It is defined in the ontology, but the datatypes of the information are not. In our approach, the entities representing data are *bt12:InformationObject* individuals. These entities are the only ones that use datatypes in a computational sense. The Computation View describes the functional decomposition of the system [10]. A functional decomposition was performed according to the general information cycle for any collaboration [13]. Five generic components has been defined: the data storage, the data interpreter (processing the data according to the formalized knowledge), the data mapper (maps the information to the knowledge of specific actor), the planner (decide the plan to be followed), the execution controller (assist the actors in the plan execution), and the execution listener (obtains new data after the plan execution). Additionally, some support components an execution engine, the reasoner, and the user interfaces are defined. The Engineering View describes the platform-specific infrastructure that supports the computational view. In our implementation, the system was deployed in four devices, the self-care device, the health professional device, a node including most of the computational components, and a node including the data storage. All are connected by Internet Protocol version 4 and HTTP. Finally, the Technology View describes the implementation of the system in terms of a configuration of technology objects representing the hardware and software components of the implementation [10]. In this view, the technologies used to implement the functional components are selected. For the execution engine functional component, the CamundaBPMN version 7.3 was selected. For the reasoner component the SPIN Rule Reasoner in its version 1.4 was deployed, and for the data storage the Virtuoso Openlink Server version 6.1 was used. The other functional components have been implemented using the Spring Framework version 4.1.7, available as open-source projects [15].

For evaluating the proposed system, a quantitative formal experiment was performed. Effectiveness is quantified using precision, recall and F-measure metrics. The hypothesis of the experiment is that the F-measure of the system's diagnosis is higher than 0.71, it is the average of the algorithms C4.5 and CART evaluated for the diagnosis of diabetes [16]. The specialist provided 20 anonymized medical records including its observations and decisions made for these patients. The medical records were structured according with the proposed information model, and then 1200 medical observations were obtained [15]. The prototype developed operates as an expert system for glycaemic control. after entering the 20 medical records, the system provides diagnostic suggestions. Table 1 compares the outcomes for one medical record. Some diagnoses corresponding to non-diabetic complications (e.g. chondromalacia of patella, mild malnutrition) were not included in the system. Other diagnoses have not been asserted due to the difficulty to infer them using the aforementioned rules, such as no chronic complications, uncomplicated diverticular disease colon, or probable primary hypothyroidism. Some

correct states about the patient have been added by the designed system (bolded in Table 1). These diagnoses were considered irrelevant by the specialist because they are not included in ICD10, which is used to classify the relevant medical diagnoses in Colombia. The mean F-measure obtained was 0.89, with a minimum value of 0.57 and a maximum value of 1. The F-measure has a standard deviation of 0.13. The mean precision is 1 and recall is 0.82. The significance of the results was evaluated with a one-sample T-test. The one-sample T-test resulted in an F-measure significantly higher than the threshold value (0.71) with a value of $p=0.00$.

Table 1. Diagnoses by physicians versus diagnoses by the system (underlined and bolded the differences)

Diagnosis by Physician	Diagnosis by the System
Type 2 diabetes mellitus	Type 2 diabetes mellitus
Peripheral diabetic neuropathy	Peripheral diabetic neuropathy
Overweight	Overweight
Metabolic syndrome	Metabolic syndrome
Hypertriglyceridemia	Hypertriglyceridemia
Raised fasting plasma glucose	Raised fasting plasma glucose
<u>Scleral and hypertensive cardiopathy</u>	South Asian central obesity
<u>Congestive heart failure stage II – C</u>	Decreased ankle reflex
<u>Coronary artery disease</u>	Medical alert – hyperglycemia
	Hypoesthesia
	Medical alert – hypertension

4. Discussion

The correct and formal description of domains and contexts through rules allows the adaptability of the system. For example, the presented system was developed using a description of the context of the Colombian health system, but it can be adapted to other health systems by the definition of their specific context. Furthermore, the software system is able to adapt to the physician's preferred language and methodologies including internal guidelines, measurement units, etc.

Contrary to the presented proposal, traditionally, the development team models the system by abstracting the types of information generated and shared during the business process. The resulting models are semi-formal descriptions of the system and are not intended to formally describe the system domains' knowledge, thereby preventing knowledge sharing and management including the deployment of best-of-breed expertise. Therefore, at least the following problems arise in traditional development processes:

1. The models are highly dependent on the knowledge of the development team. Different development teams provide heterogeneous models without a clear way of harmonization.
2. The models ignore essential concepts and knowledge of the business domain, as domain experts are usually not part of the team.
3. The models cannot guarantee correct inferences using logic rules.
4. Most parts of the models are specific for the correspondent business process, limiting the re-usability of components and reducing the chance of interoperability.
5. There is no a clear separation between the business domain description and the description of the information objects. That makes interoperability between information models difficult.

One limitation of the presented proposal is the medically relevant diagnoses of absence conditions, e.g. the diagnoses “No chronic complications” and “Uncomplicated diverticular disease colon”.

5. Conclusions

This paper presents the development of a software system and its evaluation process, starting from an abstract description of the real system considering computation-independent aspects up to a concrete ontology-based web application. This application follows medical guidelines, includes the data of the patients and suggests a precise diagnosis.

The paper testifies the feasibility to develop an adaptive, interoperable and intelligent T2DM Care System based on the GCM methodology and framework. The resultant software is available as open source in [15].

References

- [1] Blobel, Bernd, "Challenges and Solutions for Designing and Managing pHealth Ecosystems", *Frontiers in medicine*, vol. 6, article 83, 2019.
- [2] Blobel, B. and Oemig, F., "Solving the modeling dilemma as a foundation for interoperability", *European Journal for Biomedical Informatics*, vol. 14, no. 3, pp.3-12, 2018.
- [3] Blobel, Bernd, and Frank Oemig. "The importance of architectures for interoperability", *Stud Health Technol Inform*, vol. 211, pp. 18–56, 2015.
- [4] Health Level 7 International, "HL7 Version 3 Standard: Security and Privacy Ontology, Release 1.0," *Ann Arbor: HL7 International*, 2013.
- [5] "Diabetes", *World Health Organization*, 2015. [Online]. Available: <http://www.who.int/news-room/fact-sheets/detail/diabetes>. [Accessed: 03-Aug-2018].
- [6] S. Schulz and M. Boeker, "BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application", presented at the Workshop on Ontologies and Data in Life Sciences, Koblenz, Germany, 2013, pp. 19–20.
- [7] B. Smith *et al.*, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration", *Nature biotechnology*, vol. 25, no. 11, pp. 1251–1255, 2007.
- [8] International Organization for Standardization, "ISO/IEC 19510:2013 Information technology - Object Management Group Business Process Model and Notation." Geneva:ISO, Nov-2013.
- [9] H. Knublauch, J. A. Hendler, and K. Idehen, "SPIN-overview and motivation", *W3C Member Submission*, 2011. [Online]. Available: <http://www.w3.org/Submission/spin-overview/>. [Accessed: 09-Jan-2017].
- [10] ISO, "Information technology — Open Distributed Processing — Reference model: Overview", Nov-2008. [Online]. Available: <http://www.itu.int/rec/T-REC-X.901-199708-I/en>. [Accessed: 09-Jan-2017].
- [11] G. A. Uribe, B. Blobel, D. M. López, and S. Schulz, "A generic architecture for an adaptive, interoperable and intelligent type 2 diabetes mellitus care system", *Stud Health Technol Inform*, vol. 211, pp. 121–131, 2015.
- [12] G. A. Uribe, B. Blobel, D. M. López, and A. A. Ruiz, "Specializing architectures for the type 2 diabetes mellitus care use cases with a focus on process management", *Stud Health Technol Inform*, vol. 211, pp. 132–142, 2015.
- [13] G. A. Uribe, B. Blobel and D. M. López, "Systems architectures and ontologies applied to health information systems", 1st ed., Editorial Universidad del Cauca, Popayán, 2019.
- [14] International Labour Organization, "International Standard Classification of Occupations (ISCO)", *ISCO - International Standard Classification of Occupations*, 2007. [Online]. Available: <http://www.ilo.org/public/english/bureau/stat/isco/>. [Accessed: 06-Apr-2014].
- [15] G. Uribe, "Type 2 diabetes mellitus Project", 20-Sep-2018. [Online]. Available: <https://github.com/gaurgo/DiabetesCare>. [Accessed: 10-Sep-2019].
- [16] J. S. Giboney, S. A. Brown, P. B. Lowry, et. al., "User acceptance of knowledge-based system recommendations: Explanations, arguments, and fit", *Decision Support Systems*, vol. 72, pp. 1–10, 2015.

Optimization of Entropy-Based Automated Dyslalia Screening Algorithm

Emilian Erman MAHMUT^{a,1}, Dorin BERIAN^a, Michele DELLA VENTURA^b and Vasile STOICU TIVADAR^a

^aPolitehnica University Timisoara, Romania

^bDepartment of Music Technology, Music Academy “Studio Musica”, Treviso, Italy

Abstract. This paper presents the current state of progress of a project aimed at achieving an automated information entropy-based discrimination of phoneme mispronunciations in utterances of early school-age children. The introductory part briefly describes the dyslalia symptomology and the incidence of dyslalic disorders. This section also reviews the current challenges posed by the main research objective in other similar projects sharing the same objective and summarizes the current results thereof. The Material and Method section presents the conditions, the technology and the feature-extraction technique used in the experiment. The same section also describes the computation of the information entropy values of each analyzed speech sample. The highest match rate of 93.33% was achieved in the classification of words containing the phoneme /r/ in the initial position. A synthesis of the achieved results is provided in the Results section based on which conclusions are drawn and exposed in the Discussion and Conclusions section.

Keywords. Dyslalia, Information Entropy, Polynomial Trendline

1. Background

Dyslalia is the most frequent Speech Sound Disorder (SSD) among early-school aged children affecting mainly the phonetic tier of communication and it consists in the mispronunciation (distortion, omission, inversion or substitution) of speech sounds (phonemes). Its prevalence indicator for Romanian children is 13% [1] and its incidence is on a global ascending trend. Unless diagnosed and treated in due time, dyslalic disorders may have a dramatic impact on children's school performance and even lead to behavioral disorders. From a neuroscience perspective dyslalia has been linked to right-brain dominance and left-handedness and roughly a quarter of dyslalic subjects are affected by dyslexia-dysgraphia [2].

Phonological assimilation, which is both progressive and regressive within an utterance, poses a major challenge to the attempts to devise mathematical models for phoneme classification. It follows that the phonetic context of a given target consonant is of utmost importance for rigorous analysis of its articulation.

There are several projects aiming to provide a computerized solution for the discrimination of misarticulated phonemes. Logomon [3] is a complex speech therapy

¹ Emilian-Erman MAHMUT, Politehnica University Timisoara, P-ta Victoriei no. 2, Timisoara, Romania; E-mail: emilian.mahmut@aut.upt.ro.

software for Romanian equipped with a fuzzy-logic based expert system dedicated to pronunciation assessment. No specific pronunciation assessment results are provided by the authors, but human intervention (phoneme scores assigned by the Speech Therapists) is massively required to achieve such objective. Intelligent System for Impaired Speech Evaluation [4] focuses on the pronunciation of consonant /r/ in Romanian in an attempt to generate real-time numerical information on mispronunciations by analyzing Mel-Cepstral coefficients extracted from phonemes and by computing, for each item of the unclassified classes (test data), the (Euclidian) distance to each item in the known (learning) data classes (manually selected correct or incorrect pronunciations). The related distances are ordered and the final decision is made by using a K-Nearest Neighbor (KNN) algorithm. This study reports a 78% match rate for a K of 11. Self-Organizing Maps for Identifying Impaired Speech [5] describes an automated method for the discrimination of heavily impaired pronunciations of /r/ in Romanian based on Kohonen neural networks. The algorithm uses the parameters extracted from the alternating component of the signal envelope as feature vectors in a neural network-based classification stage. The study reports an 82.5% match rate with the best mispronunciation classification results within the 1000-1500 training ages range and concludes that mispronounced speech sounds have a tendency to organize on a single neuron while correct pronunciations spread on several map nodes. Vocaliza [6] is a computer-based speech therapy application for Spanish speaking children. Its phoneme mispronunciation module proposes a detection method based on feeding single phonemes (segmented from words uttered by children) to a phonetic decoder (automatic speech recognition algorithm). The speech sound discrimination scenario consists of 3 stages: identifying the phoneme-level confidence measure by using posterior probabilities resulting from a confusion network (CN) generated from the phone lattices (lattice compression algorithm), adapting the acoustic model by using Hidden Markov Models (HMM) trained on a Spanish language audio-record database containing 63193 words (Albayzin) and non-linear mapping of CN posteriors (confidence measures) to improve prediction as compared to the classification performed by non-expert human labelers. The authors conclude that the decision made by an expert (speech pathologist) will invariably have a subjective component. The study reports a match rate of 85.81% for the human labeling tasks. Human intervention is still required to a significant extent in order to reach efficient SSD assessment match rates enabling rigorous detection of phoneme mispronunciations.

The main objective of our research project is the development of a software solution for the fully automated entropy-based screening of dyslalic disorders in early school-age children. The screening system architecture is presented in [7].

2. Material and Method

An observational experiment was conducted on a target population of 30 children aged 5-7 from the Banatean National College in Timisoara (Romania), focusing on the pronunciation of the vibrant lateral consonant /r/ in initial, median and final position within the Romanian words: "RAFT" (shelf), "PARĂ" (pear) and "FAR" (headlight). The audio samples were recorded in the speech therapy practice using an Olympus LSP1 Linear PCM Recorder (PCM 44.1 kHz/16 bit). Each pronunciation was assessed by the speech language pathologist (SLP) so as to be compared with the results of the

screening application. The application was built in C# (Microsoft.NET) using Accord.NET audio processing libraries to display the polynomial trendline based on the positive amplitude values of the audio signal (Figure 1).

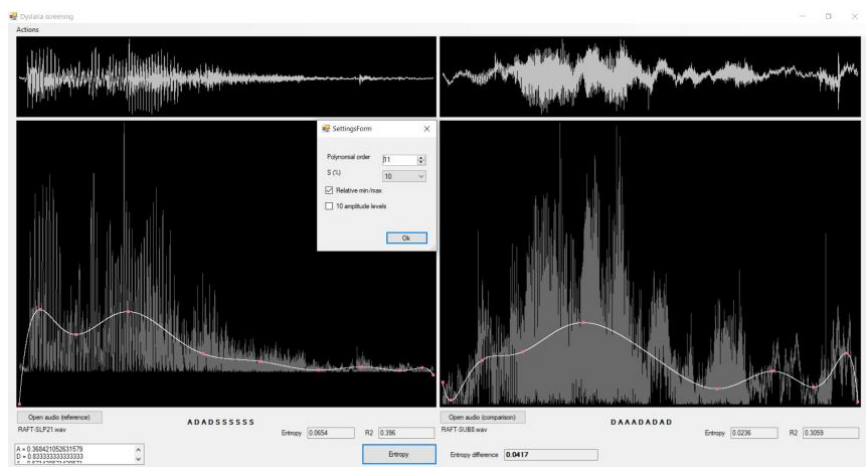


Figure 1. Dyslalia Screening Application

The algorithm extracts the polynomial trendline of a variable order (from 9 to 12) from the audio signal and describes its evolution by assigning 3 possible values to its peaks and troughs: A (ascending), D (descending) or S (stable, i.e. curve variation from previous value remains within a predefined range, in percentages of the max. amplitude value). A transition matrix shown in Figure 2 (Markov process) [8] is then filled in with such values (letters) for the reference signal (SLP’s pronunciation whose corresponding polynomial trendline has the highest R-squared value) and the test signal (subject’s pronunciation of the same word) taken together, e.g.: SLP’s pronunciation and Subject’s pronunciation. The probability of each alphabet letter is given by the quotient of the number of occurrences of the specific letter $n+1$ given another previous letter n (numerator) and the total number of occurrences of the specific letter in the 2 segments combined (denominator) [9]. The corresponding information entropy values for each segment are then calculated based on the probabilities provided by the transition matrix and compared. To have a drastic assessment of the test signal, the similarity threshold was set to a maximum value of 0.00099, therefore the test-signal entropy values belonging to the [0-0.00099] range were deemed to match the reference-signal entropy value, whereas test-signal entropy values of higher magnitude orders (i.e. in the [0.001 – 1] range) were labeled as mismatch (false positive or false negative cases).

	A	D	S		Alphabet	Total
A	2	5	0	A	7	
D	4	0	1	D	6	
S	0	1	5	S	6	
					19	

Figure 2. Transition matrix and alphabet

3. Results

Calculations were made based on the algorithm described above for 1440 values in all possible configurations (combinations of polynomial orders and S-letter ranges).

Table 1. Highest match rates for initial /r/

Initial /r/	Poly 9-S 15	Poly 10-S 2	Poly 10-S 5	Poly 11-S 10	Poly 12-S 15	SLP Opinion
Subject 1	-0.0099	-0.0614	-0.0125	0.0163	0.0049	Pararhotacism (/iaf-/)
Subject 2	0.0043	-0.0194	-0.0194	0.0102	0.0019	Rhotacism (/Raft/)
Subject 3	0.0057	-0.0617	-0.0703	-0.0218	-0.0137	Polymorphic (/rarf/)
...						
Subject 30	0.0019	-0.1449	-0.018	0.0341	0.0086	Rhotacism (/Ra:ft/)
Match Rate	86.66%	90%	90%	93.33%	86.66%	

The resulting match rate was compared to the SLP’s opinion on each analyzed utterance. The highest match rates achieved were: 93.3% for initial /r/, 80% for median /r/ and 83.3% for final /r/ as shown in Figures 3 and 4.

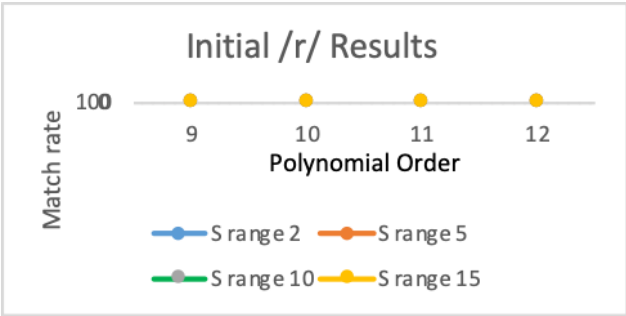


Figure 3. Initial /r/ results.

The highest match rate for the pronunciation containing the phoneme /r/ in the initial position (93.33%) was obtained using a polynomial order of 11 and an S-letter range of 10, while the best results for median /r/ (80%) were achieved using a polynomial order of 10 and an S-letter range of 5.

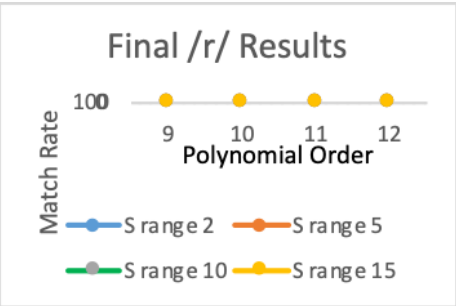
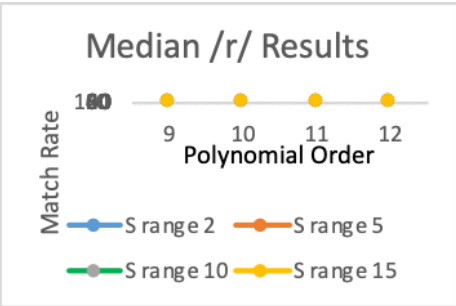


Figure 4. a) Median /r/ results b) Final /r/ results

For final /r/ the highest match rate (83.33%) was obtained by setting the polynomial order to 10 and the S-letter range to 10 or 15. It can be inferred from the synthetic diagrams (Figures 3 and 4) that an S-letter range of 2 produces the most stable results across the polynomial order range [9-12] in all three cases. The S-letter range of 15 generates the largest variation in the initial and final /r/ cases, while it remains fairly stable for the median /r/ case.

4. Discussion and Conclusions

The application computes and stores the R-squared value for each polynomial trendline to measure the goodness of fit of the model. Further analysis of such stored values is required to establish useful correlations. Higher polynomial orders (12 and above) generate poor match rates (33.33% to 70%) across the 3 cases, i.e. initial, median and final /r/.

The algorithm produced better results than previous studies and requires little human intervention.

Future work will focus on optimizing the audio signal quality (noise filtering) and segmentation by performing a normalized correlation of the pair of analyzed signals and using logatomes (monosyllabic pseudowords) with specific structures in terms of vowels (V) and consonants (C): VCV/CVC/CVV/VVC, where C stands for the target consonant. The inflection points of the polynomial trendline will also be considered for assignment of additional letters or sub-letters to be used in the transition matrix, by computation of the first and second derivatives.

Formant frequencies of the analyzed signals will also be explored to assess the possibility of devising alternative means to validate the segmentation described above.

References

- [1] E. Verza, *Tratat de logopedie*, Editura Semne, Bucuresti, Romania 2000.
- [2] J. Law, J. Boyle, F. Harris, A. Harkness and C. Nye, Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature, *International Journal of Language and Communication Disorders* 2 (2000).
- [3] S.Gh. Pentiuc, O. A. Schipor, M. Danubianu, D.M. Schipor, Automatic Recognition of Dyslalia Affecting Pre-scholars, arXiv:1406.0495v1 [cs.CY] (2014).
- [4] O. Grigore, C. Grigore, V. Velican, Intelligent System for Impaired Speech Evaluation, Recent Advances in Circuits, Systems and Signals, Book Series: International Conference on Circuits Systems Signals (2010).
- [6] O. Grigore and V. Velican, Self-Organizing Maps For Identifying Impaired Speech, *Advances in Electrical and Computer Engineering (AECE)* 3 (2011).
- [7] S.C. Yin, R. Rose, O. Saz and E. Lleida, A study of pronunciation verification in a Speech Therapy Application, *IEEE International Conference on Acoustics, Speech and Signal Processing* (2009).
- [10] E.E. Mahmut, S. Nicola, V. Stoicu-Tivadar, A Computer-based Speech Sound Disorder Screening System Architecture, *Proceedings of the 17th International Conference on Informatics, Management and Technology in Healthcare (ICIMTH)* (2018).
- [11] S. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*, Second Edition, Cambridge University Press, 2009.
- [12] M. Della Ventura, The Influence of the Rhythm with the Pitch on Melodic Segmentation, *Proceedings of the Second Euro-China Conference on Intelligent Data Analysis and Applications* (Springer) (2015).

Parallel Sentence Alignment from Biomedical Comparable Corpora

Rémi CARDON^a and Natalia GRABAR^a

^aUMR CNRS 8163 – STL, F-59000 Lille, France

Abstract. Parallel sentences provide semantically similar information which can vary on a given dimension, such as language or register. Parallel sentences with register variation (like expert and non-expert documents) can be exploited for the automatic text simplification. The aim of automatic text simplification is to better access and understand a given information. In the biomedical field, simplification may permit patients to understand medical and health texts. Yet, there is currently no such available resources. We propose to exploit comparable corpora which are distinguished by their registers (specialized and simplified versions) to detect and align parallel sentences. These corpora are in French and are related to the biomedical area. We treat this task as binary classification (alignment/non-alignment). Our results show that the method we present here can be used to automatically generate a corpus of parallel sentences from our comparable corpus.

Keywords. sentence alignment, text simplification, classification

1. Introduction

Parallel sentences provide semantically similar information which can vary on a given dimension. The dimension on which the parallelism is positioned can come from many levels, here we are concerned with expert and non-expert register of language. The following pair of sentences illustrates this:

- Expert : *Drugs that inhibit the peristalsis are contraindicated in that situation*
- Non-expert : *In that case, do not take drugs intended for blocking or slowing down the intestinal transit*

Pairs differentiated by their degree of technicality can be used for text simplification. The purpose of text simplification is to provide simplified versions of texts, in order to remove or replace difficult words or information.

Automatic text simplification can be used as a preprocessing step for NLP applications or for producing suitable versions of texts for humans. In this second case, simplified documents are typically created for children [1], or for people with mental or neurodegenerative disorders [2].

Helping patients to better understand medical and health information is an important issue, which motivates our work[3].

In order to perform biomedical text simplification, we propose to collect parallel sentences, which align difficult and simple information. We can exploit an existing monolingual comparable corpus with medical documents in French [4]. The corpus is composed

of documents created for medical professionals and documents created for patients. The purpose of our work is to detect and align parallel sentences from this comparable corpus. We also propose to test what the impact of imbalance on categorization results is: imbalance of categories is indeed a natural characteristic in textual data.

The existing work on searching parallel sentences in monolingual comparable corpora indicates that the main difficulty is that such sentences may show low lexical overlap but be nevertheless parallel. This task is usually explored in general-language corpora and performed as assigning a similarity score from 0 to 5[5]. Among the exploited methods, we can notice lexicon-based methods which rely on similarity of subwords or words from the processed texts or on machine translation [6]; knowledge-based methods which exploit external resources, such as WordNet [7]; syntax-based methods which exploit the syntactic modelling of sentences. [8]; or corpus-based methods [9]. There is no existing work on building a corpus for text simplification in the biomedical domain.

2. Method

We use the CLEAR comparable medical corpus [4] available online¹ which contains three comparable sub-corpora in French. Documents within these sub-corpora are contrasted by the degree of technicality of the information they contain with typically specialized and simplified versions of a given text. These corpora cover three genres: drug information, summaries of scientific articles, and encyclopedia articles. The *Drugs* corpus contains drug information such as provided to health professionals and patients. This corpus is built from the public drug database² of the French Health ministry. The *Scientific* corpus contains summaries of meta-reviews of high evidence health-related articles, such as proposed by the Cochrane collaboration and the simplified versions that they provide[10]. This corpus has been built from the online library of the Cochrane collaboration³. The *Encyclopedia* corpus contains encyclopedia articles from Wikipedia⁴ and Vikidia⁵. Wikipedia articles are considered as technical texts while Vikidia articles are considered as their simplified versions (they are created for children from 8 to 13 year old). Only articles indexed in the medical portal are exploited in this work.

We exploit a reference dataset with sentences manually aligned by two annotators within this corpus.

2.1. Automatic Detection and Alignment of Parallel Sentences

Automatic detection and alignment of parallel sentences is the main step of our work. The objective is to assess whether two given sentences have the same meaning.

The reference data with aligned sentence pairs, which associate technical and simplified contents, are created manually. We have randomly selected 39 documents from the corpus. The sentence alignment is done by two annotators. This alignment process provides a set of 238 equivalent sentence pairs. The inter-annotator agreement is 0.76

¹<http://natalia.grabar.free.fr/resources.php#clear>

²<http://base-donnees-publique.medicaments.gouv.fr/>

³<http://www.cochranelibrary.com/>

⁴<https://fr.wikipedia.org>

⁵<https://fr.vikidia.org>

[11], before consensus. In order to perform the automatic categorization, we also need negative examples, which are obtained by randomly pairing all sentences from all the document pairs except the sentence pairs that are already found to be parallel. Approximately 590,000 non-parallel sentences pairs are created in this way. That high degree of imbalance is the main challenge in our work and we address it in the experimental design (sec. 2.2).

For the automatic alignment of parallel sentences, we use a binary classification model that relies on the random forest algorithm [12]. The implementation we use is the one that is available in scikit-learn [13]. The features we use are the following: number of common non-stopwords; percentage of words from one sentence included in the other sentence, computed in both directions; sentence length ratio; average word length ratio, total number of common character bigrams and trigrams, word-based similarity measure exploits three scores (cosine, Dice and Jaccard) character-based minimal edit distance, word-based minimal edit distance [14], syntactic token-based minimal edit distance; number of common CUI (we generate all the possible subsequences of words in both sentences and check whether they have a corresponding CUI in the UMLS/SNOMED lexicon [16]. We count the number of times that a same CUI appears in both sentences); WAVG and CWASA [17].

2.2. Experimental Design

We work on sentences that are equivalent (they both express the same meaning). 238 equivalent pairs were obtained.

We performed two sets of experiments:

1. We train and test the model with balanced data (we randomly select as many non-aligned pairs as aligned pairs), and then we progressively increase the number of non-aligned pairs until we reach a ratio of 3000:1, which is close to the real data ($\sim 4000:1$).
2. Then, for each ratio, we apply the obtained model to the whole dataset and evaluate the results. Note that the training data is included in the whole dataset, we proceed this way because of the low volume of available data.

This means that each model is evaluated twice : once on the test set, and once on the whole dataset. We only report scores for the aligned category. The results that are presented correspond to the mean values over the fifty runs. Finally, we apply the best model on another 30 randomly selected documents and manually evaluate the output.

3. Results

We present the results in Figures 1 and 2: The x axis represents the growing of imbalance (the first position is 1 and corresponds to balanced data), while the y axis represents the values of Precision, Recall and F-measure. Figure 1 presents the results for the experiments where we evaluate the model on the test set, Figure 2 presents the results for the experiments where we evaluate the model on the whole dataset.

We can observe that the use of balanced data provides very high results, both for Precision and Recall, which are very close to the reference data (> 0.90 performance). These good results in an artificial setting cannot be applied to the real dataset, as is indicated by the starting point in Figure 2

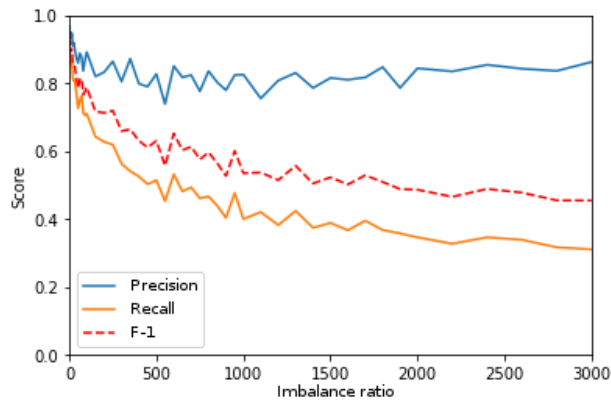


Figure 1. Evaluation on the test set.

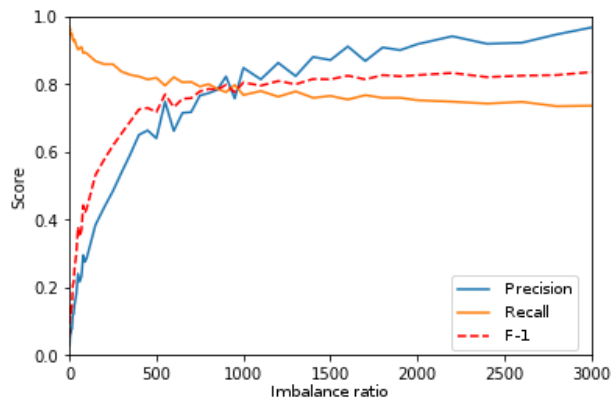


Figure 2. Evaluation on the whole dataset.

4. Discussion

When the model is learned on a substantial degree of imbalance, the Precision score is high when that model is applied to the real data, which has a ratio of about 4,000:1. The recall value is also high, but since two thirds of the aligned sentences have been used for training, that good score should be considered cautiously.

For further evaluation, we randomly selected 30 pairs of documents to evaluate the performances of the models. We used the model that was trained at a ratio of 1200:1. In terms of precision, the model shows 98.75% on all the sentence pairs aligned (80 sentence pairs), including equivalence, inclusions and intersection. Only one result showed two unrelated sentences. Those results show that we have a model that can be used to automatically generate a parallel corpus with reduced noise, from highly imbalanced comparable corpora, for text simplification purposes.

5. Conclusion

We addressed the task of detection and alignment of parallel sentences from a monolingual comparable French corpus. We use the CLEAR corpus, that is related to the biomedical area.

We made observations on the effect of imbalance during training on the performance on the real data. We show that increasing the imbalance during training increases the Precision of the model while still maintaining a stable value for Recall.

We will use that model to generate a corpus of parallel sentences in order to work on the development of methods for biomedical text simplification in French.

References

- [1] Tu Thanh Vu, Giang Binh Tran, and Son Bao Pham. Learning to simplify children stories with limited data. In LNCS 8397 Springer, editor, *Intelligent Information and Database Systems*, pages 31–41, 2014.
- [2] Ping Chen, John Rochford, David N. Kennedy, Soussan Djamshidi, Peter Fay, and Will Scott. Automatic text simplification for people with intellectual disabilities. In *AIST*, pages 1–9, 2016.
- [3] AMA. Health literacy: report of the council on scientific affairs. Ad hoc committee on health literacy for the council on scientific affairs, American Medical Association. *JAMA*, 281(6):552–7, 1999.
- [4] Natalia Grabar and Rémi Cardon. CLEAR – Simple Corpus for Medical French. In *Workshop on Automatic Text Adaption (ATA)*, pages 1–11, 2018.
- [5] Eneko Agirre, Carmen Banea, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Rada Mihalcea, German Rigau, and Janyce Wiebe. SemEval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation. In *SemEval 2016*, pages 497–511, 2016.
- [6] Nitin Madnani, Joel Tetreault, and Martin Chodorow. Re-examining machine translation metrics for paraphrase identification. In *NAACL-HLT*, pages 182–190, 2012.
- [7] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. Introduction to wordnet: An on-line lexical database. Technical report, WordNet, 1993.
- [8] Masashi Tsubaki, Kevin Duh, Masashi Shimbo, and Yuji Matsumoto. Non-linear similarity learning for compositionality. In *AAAI Conference on Artificial Intelligence*, pages 2828–2834, 2016.
- [9] Jonas Mueller and Aditya Thyagarajan. Siamese recurrent architectures for learning sentence similarity. In *AAAI Conference on Artificial Intelligence*, pages 2786–2792, 2016.
- [10] David L. Sackett, William M. C. Rosenberg, Jeffrey A. MuirGray, R. Brian Haynes, and W. Scott Richardson. Evidence based medicine: what it is and what it isn't. *BMJ*, 312(7023):71–2, 1996.
- [11] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.
- [12] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [14] Vladimir I. Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. *Soviet physics. Doklady*, 707(10), 1966.
- [15] Thupahi Sisira De Silva, Don MacDonald, Grace Paterson, Khokan C. Sikdar, and Bonnie Cochrane. Systematized nomenclature of medicine clinical terms (snomed ct) to represent computed tomography procedures. *Comput. Methods Prog. Biomed.*, 101(3):324–329, March 2011.
- [16] Sanja Stajner, Marc Franco-Salvador, Simone Paolo Ponzetto, and Paolo Rosso. Cats: A tool for customised alignment of text simplification corpora. In *Proceedings of the 11th Language Resources and Evaluation Conference, LREC 2018, Miyazaki, Japan, May 7-12, 2018*.

Performances of a Solution to Semi-Automatically Fill eCRF with Data from the Electronic Health Record: Protocol for a Prospective Individual Participant Data Meta-Analysis

Nicolas GRIFFON^{a,b,1}, Helena PEREIRA^c, Juliette DJADI-PRAT^c, María Teresa GARCÍA^d, Sara TESTONI^e, Manon CARIOU^f, Jacques HILBEY^b, Aurèle N'DJA^f, Grégory NAVARRO^g, Nicola GENTILI^c, Oriana NANNI^c, Massimo RAINERI^g, Gilles CHATELLIER^c, Agustín GÓMEZ DE LA CAMARA^{d,h}, Martine LEWI^g, Mats SUNDGRENⁱ, Christel DANIEL^{a,b}, Almenia GARVEY^j, Marija TODOROVIC^g, Nadir AMMOUR^f

^aWIND Department, AP-HP, Paris, France; ^bLIMICS, U1142, Inserm, France; ^cHEGP, AP-HP, France; ^dHospital 12 de Octubre, Madrid, Spain; ^eIRST-IRCCS, Meldola, Italy; ^fSanofi R&D, Paris, France; ^gJanssen Pharmaceutica NV, Beerse, Belgium; ^hCIBERESP; ⁱAstraZeneca, Göteborg, Sweden; ^jIcon PLC, Paris, France;

Abstract. Clinical trial data collection still relies on a manual entry from information available in the medical record. This process introduces delay and error risk. Automating data transfer from Electronic Health Record (EHR) to Electronic Data Capture (EDC) system, under investigators' supervision, would gracefully solve these issues. The present paper describes the design of the evaluation of a technology allowing EHR to act as eSource for clinical trials. As part of the EHR2EDC project, for 6 ongoing clinical trials, running at 3 hospitals, a parallel semi-automated data collection using such technology will be conducted focusing on a limited scope of data (demographic data, local laboratory results, concomitant medication and vital signs). The evaluation protocol consists in an individual participant data prospective meta-analysis comparing regular clinical trial data collection to the semi-automated one. The main outcome is the proportion of data correctly entered. Data quality and associated workload for hospital staff will be compared as secondary outcomes. Results should be available in 2020.

Keywords. Data collection; Health Information Interoperability; Clinical trial as topic; Clinical Trial Protocols as Topic; eSource.

1. Introduction

Pharmaceutical industry is facing increasing costs for drug development [1]. The automatic transfer of data from Electronic Health Records (EHR) to Electronic Data

¹ Corresponding Author, Nicolas Griffon, Web Innovation Données, 33 Bd Picpus – 75012 Paris, France; E-mail: nicolas.griffon@aphp.fr

Capture System (EDC) has been identified as one partial solution to address this issue by reducing the burden of data entry and associated activities, like data monitoring and review by sponsor [2]. Moreover, such solution would drastically reduce the time to entry, which is still too high, even with incentive [3].

A recent literature review [4] identified multiple initiatives towards automated transfer from EHR to EDC systems. Most of these initiatives being monocentric, retrospective or based on only one EHR, the authors also stress out the need for further research to better evaluate the impact of eSource solutions on data quality and workload: some burdens might decrease (entry, quality assessment...) but some others will increase (application maintenance, semantic interoperability maintenance).

The EHR2EDC project [5] has received funding from European Institute of Innovation and Technology (EIT) Health. This project is led by Sanofi R&D, include in a consortium 3 other pharma companies (see Tab 1), one Clinical Research Organization (ICON), one health data technology company (InSite, a TriNetX company), 4 hospitals (see Tab 1), the French National Institute for Medical Research (Inserm) and the European Institute for Innovation through Health Data (i~HD, a not for profit organization). The aim of this project was making such transfer a reality, by overcoming technical, organizational and regulatory difficulties. Building on the already existing InSite solution (developed by InSite), that allows feasibility studies and provides support for recruitment, a new module was developed, that prefills the concomitant medication, local laboratory, vital signs and demographic sections of the eCRF, and sends this information, under the investigator control, to the trial sponsor EDC. After unit, component and usability testing, it is now necessary to evaluate this tool's performances (data transfer capacity, error rate in comparison with manual entry, acceptability with regard to process evolution). This work focusses on the design of the evaluation protocol allowing to assess the performance of an eSource solution and the application of this evaluation protocol in the EHR2EDC context within the TransFAIR study.

2. Methods

2.1. Individual participant data – prospective meta-analysis (IPD–PMA)

IPD meta-analysis is recognized by the Cochrane collaborative group as a 'gold standard' of systematic review [6], and some authors strongly advocate for prospective meta-analysis [7]. Even if this study is a proof-of-concept technology evaluation study, and not a clinical trial, it was decided to design it like an IPD–PMA, following as much as possible PRISMA-P [8][9] and MOOSE [10] statements. This design will ease later inclusion of data from other tests in other hospitals using an eSource solution.

A preliminary protocol was developed by Sanofi R&D and submitted for review and validation to the whole EHR2EDC consortium. Each hospital was allowed to modify the template in order to customize this protocol but agreed to share patient level data to the statistician in charge of the meta-analysis (not the data collected in itself, but rather the result of the comparison of paired data sets collected, either manually or semi-automatically).

2.2. Study design

The main idea of this evaluation is to take advantage of real ongoing clinical trials (support CT), by performing a semi-automatic data collection in addition, and independently, to the usual manual data collection (see Fig 1). Support CT will be conducted in a completely usual way, and will not be affected by the TransFAIR study. The data collected in this way are named "Manual Data". All the patient included in the support CT are eligible for inclusion in the TransFAIR study.

In parallel of each support CT, the EHR2EDC transfer module will be installed allowing direct use of EHR data under the supervision of clinical investigators or study personal in a mirror study. As support CT are ongoing, patients may have been included before the beginning of the TransFAIR study. To increase the amount of data collected, both prospective and retrospective data will be collected using the EHR2EDC transfer module according to the following steps:

1. The study coordinator / investigator will supervise the automated data collection. Through the module interface, (s)he will review, validate and transfer the data required by the protocol to the sponsor eSource database.
2. The sponsor will then reconcile the manual and eSource database to identify discrepancies and provide them to the study coordinator / investigator.
3. For each discrepancy identified, the study coordinator / investigator will go back to the source documents of both eSource and support CTs data in order to collect the real value of the information collected.
4. To the best of his/her knowledge and skills, (s)he will try to identify error causes amongst: (i) rounding error, (ii) transcription error, (iii) wrong data entry for Manual Data, and (a) modification of information during the transfer, (b) wrong information transferred, (c) user error with the new module (wrong patient, wrong visit or wrong data selection) for the eSource data.

For data outside EHR2EDC scope (e.g. pathology, medical history), Manual Data will only be counted – and considered correct.

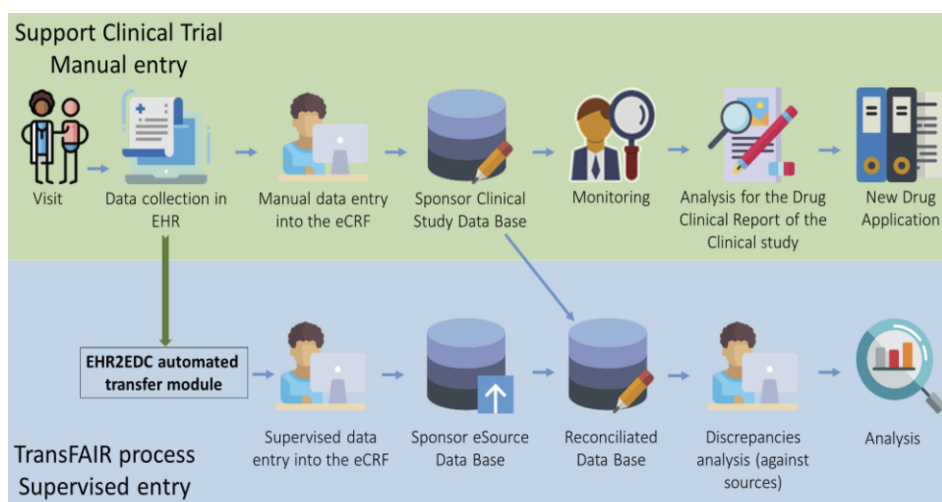


Figure 1. Support CT and TransFAIR study designs.

2.3. Outcomes

The main objective of the TransFAIR study is to assess the percentage of data accurately transferred (the number of data correctly transferred divided by the whole number of correct data – either manually or automatically entered). Secondary outcomes are:

1. Assessment of the percentage of target data points that can be accurately processed by the eSource solution. Computed like main outcome besides the denominator, that is restricted to data transferred.
2. Descriptive analysis of the identified types of inconsistency
3. Comparison of data management activities between the two processes assessed by the number of queries

Subgroup analysis will be performed on study site, type of protocol, medical specialty and data domain. For exploratory analysis, we will:

1. Compare entry error rates between arms, based on the discrepancy analysis, and try to identify some determinants of error rate: datatype (Boolean, text, date...), CDISC domain, data quality...
2. Explore more thoroughly the differences in queries between semi-automated and manual data entry.

2.4. Statistical analysis

We won't assess publication bias because of the prospective design of this meta-analysis. The heterogeneity between sites will be assessed using Cochran's Q and I square statistic. For primary outcome and 1st secondary outcome, proportion will be computed using a multi-level model to take account of inter site variance. A conditional logistic regression model will be used to compute OR estimate for entry error between arms. Once again, the inter site heterogeneity will be handled by a multi-level model.

2.5. Ethical concerns

Patients included in the support clinical trials will be asked if they consent to the reuse of their data in the context of the TransFAIR study. All the data shared between partners are shared with respect to the EU General Data Protection Regulation (GDPR).

3. Results

The TransFAIR study is being conducted using the proposed study protocol at 3 hospitals in parallel of 6 support clinical trials between mid-September and end-November, 2019 (see Tab 1). The 3 IRBs of the hospitals involved in the project approved both TransFAIR and support CT.

4. Discussion

Making the data interoperable between pharma and hospital world is a huge work that has been accomplished for demographic, local laboratory, medication and vital signs data

during the EHR2EDC project. It required the creation – to be published – of: (i) a FHIR common information model (CIM) – each hospital must be able to provide data in the CIM way, and, (ii) an important set of mappings between this CIM and pharma standards – derived from CDISC. This model makes different EHR and EDC systems interoperable.

Table 1. Description of support clinical trials

Running hospitals	Registration number ¹	Sponsor	Indication	Phase
APHP ² , 12Oct	NCT03767244	Janssen	Advanced prostate cancer	3
APHP, IRST-IRCCS	NCT03390504	Janssen	Advanced urothelial cancer	3
APHP	NCT03315143	Sanofi	Diabetes/cardiology	3
12Oct	NCT03284957	Sanofi	Advanced breast cancer	1b
12Oct	NCT03619213	Astra Zeneca	Heart failure	3b
IRST-IRCCS	NCT02516241	Astra Zeneca	Urothelial cancer	3

1: ClinicalTrials.gov identifier; 2: four APHP hospitals involved (Mondor, HEGP, Bichat, Lariboisière)

This study will allow the identification of data quality gaps in EHR, highlighting the data quality dimensions, as described by Khan et al., to put the focus on for the EHR2EDC scenario. The idea is to enhance the i-HD quality seal for Research Platforms [12] and to make it a requirement for any hospital to deploy and use any data transfer application between EHR and EDC.

This paper focused on the design of an evaluation protocol allowing the assessment of eSource solutions implementing large-scale connectivity between EHR and EDC systems. Results will be available in 2020.

References

- [1] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Economics* 47 (2016), 20-33.
- [2] Eisenstein EL, Nordo AH, Zozus MN. Using Medical Informatics to Improve Clinical Trial Operations. *Stud Health Technol Inform.* 234 (2017), 93-97.
- [3] Ahonen K, Herzezeel T, Hull J, King R, Nadolny P and Perkins R. 2001 to 2009: Investigational Site Perspectives on Clinical Trial Information Systems. *eClinical Forum* 2012. www.eclinicalforum.org
- [4] Garza M, Myneni S, Nordo A, Eisenstein EL, Hammond WE, Walden A, et al. eSource for Standardized Health Information Exchange in Clinical Research: A Systematic Review. *Stud Health Technol Inform.* 257 (2019), 115 124
- [5] EHR2EDC. Available at: <http://www.eithealth.eu/ehr2edc> [last check 30/09/2019]
- [6] About IPD meta-analyses. Available at <https://methods.cochrane.org/ipdma/about-ipd-meta-analyses> [last check 30/09/2019]
- [7] Ioannidis J. Next-generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. *Br J Sports Med.* 51(20) (2017), 1456–1458.
- [8] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 4 (2015), 1.
- [9] Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ.* 349 (2015), g7647.
- [10] Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 283 (2000), 2008–12.
- [11] Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., ... & Liaw, S. T. (2016). A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *Egms*, 4(1).
- [12] The i-HD Quality Seal for Research Platforms (QS4RP). Available at: <https://www.i-hd.eu/index.cfm/services/qs4rp/> [last check 30/09/2019]

Physician-Machine Interaction in the Decision Making Process

Saveli GOLDBERG^{a1}, Anatoly TEMKIN^b, Benjamin WEISBURD^c

^a *Radiation Oncology Department, Massachusetts General Hospital, Boston, MA, USA*

^b *Department of Computer Science, Boston University Metropolitan College, MA, USA*

^c *Broad Institute, Cambridge, MA, USA*

Abstract. We propose an approach to decision support systems (DSS) that starts with the user first making their own unassisted decision α_U and providing this decision as an input to the algorithm. Then, if the decision based of machine learning (ML) disagrees with the user's initial decision, it iteratively works with the user to converge to a common decision or at least make the user reconsider input values that are inconsistent with α_U . We provide a detailed description of this approach along with examples, and then discuss potential benefits and limitations of this approach.

Keywords. decision support; machine learning; user interaction

1. Introduction

Success of artificial intelligence systems in medicine are amazing. However, the greater the accuracy of machine learning algorithms (ML), the greater the user's exclusion from participation in these decision-making processes. The danger of this is not only in the loss of responsibility for making the decision, but also in its correctness. The point is that machine learning uses a training sample collected in the past, while decisions are made in the present. But the past may be radically different from the present and the speed of changes may exceed the speed of additional training. Moreover, the machine's decision in any case is based on the patient's data provided by the physician, and therefore there is a danger of the preliminary physician decision's influence on the accuracy of this data. In psychology, this effect is called "Cognitive bias" [1]. Even in the absence of subjectivity in the assessment of the attribute's value there could be an error in the instrument's reading, in the measurement method itself, and in the natural fluctuations of the meter's readings. The best defense against these threats (the effect of detachment and non-availability of data) is seen in automating the explanation of the machine solution and the obligatory responsibility of a physician in the final decision [2]. These measures are indeed very helpful, but justification remains challenging, particularly for complex ML models.

Additionally, the better ML systems become, the more likely users will stop putting much effort into analyzing or critically evaluating the algorithms' decisions, even if automated explanations are also provided. Here we propose an approach that

¹ Saveli Goldberg, Ph.D., Radiation Oncology Department, Room 116, Massachusetts General Hospital, 10 Emerson Pl, Boston, MA, 02114. Email: sigoldberg@mgh.harvard.edu

introduces a meta-agent or decision support system (DSS) between the user and the ML algorithm [3]. This DSS restructures the interaction between a user and the ML to mitigate the potential loss of expertise and restore a fuller sense of responsibility to users. Central to this is that it requires the user to first make an unassisted decision and provide this as an input parameter to the algorithm before the algorithm generates its own automated decision. The DSS is trying to find "weaknesses" in the decision of the user. If the decision of the user continues to differ from the decision of the ML, the DSS helps the user to find the reasons for these discrepancies.

2. The Algorithm Overview

2.1 Definitions:

Let \mathbf{X} be the set of strings of length n of numerical or Boolean values and $D = \{\alpha_j\}$, $j=1, \dots, k$, is the set of possible decisions that could be made by the user or ML. Each string represents n parameter values. We define the decision function $f: (\mathbf{X}, D) \rightarrow [0,1]$, that maps an input $\mathbf{x}=[x_1, \dots, x_n]$ and a decision α to a confidence interval $[0,1]$.

Let $\mathbf{v} = [v_1, \dots, v_n]$ be the input provided by the user and α_u is the initial unassisted decision made by the user. The user also singles out parameters $[v_1, \dots, v_m]$, $m \leq n$, as being particularly important for making the decision α_u . ML finds $\max(f(\mathbf{v}, \alpha))$ for all α that belong to D . Let α_{ml} be the ML algorithm decision based on the user-provided input $\mathbf{v} = [v_1, \dots, v_n]$. The values x_i that are entered into the ML algorithm may be biased or erroneous, so we define $\Omega(x_i)$, $i=1, \dots, n$ as the set of values which are considered within the error bounds for x_i . The bias includes the uncertainty of a patient and the physician (i.e. the user). (the level of a patient headache or the degree of paleness are examples of a possible bias). When there is an uncertainty in assessing a parameter, we have a phenomena of "confirmation bias" and "selective perception" [1].

We introduce a parameter normalization x_i^{norm} for each i , based on four thresholds: $a_{0i}, a_{1i}, a_{2i}, a_{3i}, a_{4i}$ that are defined by a group of expert physicians.

$$\begin{aligned} x_i < a_{0i} & : \text{strong deviation } x_i^{\text{norm}} = 0 + x_i / a_{0i} \\ a_{1i} < x_i < a_{2i} & : \text{abnormal: } x_i^{\text{norm}} = 1 + (x_i - a_{1i}) / (a_{2i} - a_{1i}) \\ a_{2i} < x_i < a_{3i} & : \text{normal: } x_i^{\text{norm}} = 2 + (x_i - a_{2i}) / (a_{3i} - a_{2i}) \\ a_{3i} < x_i < a_{4i} & : \text{abnormal: } x_i^{\text{norm}} = 3 + (x_i - a_{3i}) / (a_{4i} - a_{3i}) \\ a_{4i} < x_i & : \text{strong deviation: } x_i^{\text{norm}} = 3 + x_i / (a_{4i}) \end{aligned}$$

Thus, normalized parameters will belong to five intervals: $[0,1]$, $[1,2]$, $[2,3]$, $[3,4]$, and the interval of numbers greater than 4.

Using this normalization we substitute $[x_1, \dots, x_n]$ for $[x_1^{\text{norm}}, \dots, x_n^{\text{norm}}]$. Now we define the distance between strings \mathbf{x} and \mathbf{y} in a standard way as

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$$

To define a range of sub-normal (deviated) values, a team of experts empirically establishes acceptable parameter limits. They are determined for a certain combination of parameters and certain patients. If a parameter stays within the defined abnormal or normal range, no action is required. The strong deviation range covers all the sets of possible values beyond the abnormal values [4]. For example, if the body temperature is less than 35°C , then it is a strong deviation. If it is in the range $35^\circ\text{--}36^\circ$, then it is considered abnormal. If it is in the range $36.1^\circ\text{--}37.5^\circ$, then it is normal. If the range is $37.6^\circ\text{--}38.3^\circ$, then it is abnormal, and if it is more than 38.3° , then it is a strong deviation.

2.2 The Algorithm [5]:

- Step 1.** The user enters a string of data, $v = [v_1, \dots, v_n] \in X$ into the ML algorithm.
- Step 2.** The user makes an unassisted decision α_U and enters it into the ML.
- Step 3.** The user explains his or her decision α_U . The user singles out m out of n input values as being particularly important in making the decision α_U .
- Step 4.** DSS tests the physician decision. In this step the DSS checks whether α_U is stable when the input parameters are perturbed within the error bounds $\Omega(v_i)$. If ML comes up with the decision α_{ML} that is not equal to α_U , then go to Step 5. If ML comes up with the decision α_{ML} that is equal to α_U , then go to Step 7.
- Step 5.** DSS selects one or more “suspicious” parameters relied on by the user in making the initial decision. Since $\alpha_U \neq \alpha_{ml}$, DSS iteratively works with the physician. DSS could, at this point, just show the value of α_{ml} to the physician, but DSS specifically avoid doing this to prevent the physician from automatically changing his or her decision to α_{ml} . Instead, DSS use an indirect approach where it tries to find the parameter whose value v_i deviates the most from the typical value for the given α_U .
- There are some subtle details about this approach intended to make the user more willing to review his or her initial decision α_U . The First, DSS takes the user’s point of view as the starting point and works from there. The Second, DSS tries to only select parameters i that are greater than m , because the user doesn’t consider them as particularly important to α_U , so it may be more productive to point the user to inconsistency in these parameters rather than to inconsistency in parameters $1, \dots, m$ which the user marked as supporting α_U .
- Step 6.** DSS explains ML decision. As at this point the user’s decision still differs from the ML’s decision, the DSS attempts to explain the difference between the ML’s decision α_{ml} and the user’s decision α_U in a way that is intuitive for a user rather than in a way that is based on the ML’s internal representation. For this purpose, the DSS determines which input parameters were the most important for the ML’s decision. This can be done by finding the input vector z' which is the closest to the user’s input values v and which prompts the ML to change its decision from α_{ml} to α_U . A critical part of this step is that the distance between points v and z' is computed in the normalized parameter space (X^{norm}). The DSS can use a grid search to find points on the boundary between α_{ml} and α_U . Once z' is found, the parameters in which z' and v differ the most are recognized as the parameters being the most important to explain the difference between α_{ml} and α_U .
- Step 7.** A physician tests the ML decision and makes the final decision. A physician can modify v and observe corresponding decisions of ML. ML can in turn change its decision and provide an updated justification. Once the physician receives the ML decision for all the cases of interest, he or she makes the final decision.

3. Example Application

Some features of the proposed approach were implemented in the decision support system “Dinar-2” for the Center for Child Air-Ambulance Services in Yekaterinburg, Russia. One of the goals of the Center was to provide a remote consultation to doctors

in regional medical centers who treated seriously ill children. The Center has served the large geographic area, so for its air-ambulance services it would often take a long time to reach regional centers. A computerized decision support system had been key to the efficient operation of the Center. This system helped in diagnosing the type of pathology and in determining its severity. It has also assisted in selecting the best course of action and the healthcare center that is best suited for treating patients. Besides test results, the system had to consider a significant amount of subjective information about the patient's condition. This made the decision support task more complicated because the subjective information was susceptible to conscious and subconscious biases on the part of the reporting physicians. Specifically, these biases tended to skew the provided information toward making a patient's condition appear either severe than it was. Because of that, the Dinar-2 decision support system assigned an a-priori confidence interval to every input parameter that was based on subjective information. Then, the system perturbed the inputs within the bounds of these confidence intervals and checked whether its diagnosis was consistent with the diagnosis initially proposed by a physician. If, under these perturbations, Dinar-2's diagnosis of the patient's condition did not match that of the user, Dinar-2 would follow the proposed interaction flow (described in section II above) to clarify the diagnosis. After its initial deployment, several years ago, Dinar-2 was soon adopted by 39 air-ambulance centers across former Soviet Union, and has been use since then. For example, available statistics for the Yekaterinburg region for 2017 indicate that during that year the system assisted in evaluating 537 cases. In 131 of these cases (24%), effective remote diagnosis and consultation proved sufficient for resolving the patient's crisis, and the need to dispatch an air-ambulance was avoided [6].

4. Discussion

Can it be of interest to have only a user testing the machine or the machine testing a user? In our opinion, it can be. There are situations in which the degree of responsibility is so big that it is not desirable to directly indicate the machine's views. For example, this may occur when critical decisions about the patient care are to be made. Then, the parameters in which the ML and physician opinions differ are presented to the DSS for an increased scrutiny without ML being mentioned. A similar approach was used in [7]. Explaining an ML classifier's decision while treating the classifier as a black box was proposed before. However, our approach serves to provide explanations that are more intuitive to a physician because it uses the normalization of input parameters to map them into a space that is based on the way experts naturally think about these parameters. We would like to emphasize the importance of a physician to reach a conclusion before following up on a machine's suggestion. Requiring physicians to first reach their own decision serves to counteract the loss of physician's expertise and sense of responsibility that often occurs when users delegate decisions to the DSS. It prevents physicians from becoming complacent and motivates them to give more thought to their initial decision. It provides continued opportunity for users to revisit and refresh their domain knowledge. When a physician decision is different from the algorithm's, he or she reconsiders that decision considering parameters highlighted by the algorithm. In the end, it makes it more likely that the user will critically evaluate the machine's decision. Finally, if some of the inputs provided by the user are based on subjective information, it is likely that these will

contain biases that support the user's unassisted decision. By having access to this decision, the algorithm can take steps to correct these biases. Of course, all these benefits and the user-machine dialogue itself will not work if a user's solution is just a formality and a user will be waiting for a machine's solution before resorting to any actions. If the original decision is recorded, a physician will think twice before making it to avoid possible accusations of incompetency. Gaining experience of such a communication with the machine seems to significantly benefit each expert, but it also is useful for improving the quality of ML when working with an expert. Ongoing work involves testing this approach in medical diagnostic applications.

5. Conclusions

In conclusion, there are several benefits to structuring decision support systems in a way that makes the user provide their own unassisted decision to a decision support system as a first step. We expect this approach will increase the accuracy of the final decision and will serve to maintain and possibly even improve the expert users' domain knowledge.

References

- [1] Plous, Scott, *The Psychology of Judgment and Decision Making*, 1993, p. 233.
- [2] Bryce Goodman, Seth Flaxman. European Union regulations on algorithmic decision-making and a "right to explanation" *AI Magazine*, 2017, Vol 38, No 3,
- [3] Goldberg S., Katz G., Weisburd B., Belyaev A., Temkin A. Integrating User Opinion in Decision Support Systems. In: Arai K., Bhatia R. (eds) *Advances in Information and Communication. FICC 2019. Lecture Notes in Networks and Systems*, vol 70. Springer, Cham
- [4] Goldberg S, Nikita Shklovskiy-Kordi, Boris Zingerman . Time-oriented multi-image case history - way to the "disease image" analysis. *VISAPP*, 200-203, 2007
- [5] Goldberg S, Galitsky B, Weisburd B. Framework for interaction between expert users and Machine Learning Systems. 2019 http://ceur-ws.org/Vol-2448/SSS19_Paper_Upload_217.pdf
- [6] Report Neonatology Department of Sverdlovsk State Children Hospital, Russia, 2018, 43-47
- [7] Goldberg, S., Niemierko, A., Shubina, M., & Turchin, A. "Summary Page": A novel tool that reduces omitted data in research databases. *BMC Medical Research Methodology*, 2010, 10, 91-97.

Predicting Length of Stay in Hospital Using Electronic Records Available at the Time of Admission

Marta WILK^{a,d}, D William R MARSH^{a,d1}, Sarah DE FREITAS^c, John PROWLE^{b,d}

^a*School of Electronic Engineering and Computer Science*

^b*William Harvey Research Institute, Barts and the London School of Medicine and Dentistry*

^c*Acute and Renal Medicine, The Royal London Hospital, Barts Health NHS Trust*

^d*Queen Mary University of London*

Abstract. Predicting a patient's hospital length of stay (LoS) can help manage staffing. In this paper, we explore LoS prediction for a large group of patients admitted non-electively. We use information available at admission, including demographics, acute and long-term diagnoses and physiological tests results. Data were extracted from the electronic health records (EHR), so that the LoS prediction would not require additional data entry. Although the data can be accessed, the system does not present a unified view of the data for one patient: to resolve this we designed a process of cleaning and combining data for each patient. The data was used to fit semi-parametric, parametric and competing outcomes survival models. All models performed similarly, with concordance of approximately 0.7. Calibration results showed underestimation of predicted discharges for patients with high discharge probabilities and overestimation of predicted discharges for those with low discharge probabilities. The main challenges in operationalizing LoS predictions are delays in entering admissions data into EHR and absent data about non-medical factors determining discharges.

Keywords. Length of Stay, Bed management, Survival analysis, Medical data transformation.

1. Introduction

We develop and evaluate a Length of Stay (LoS) predictive model and use the prediction of individual patients' LoS from daily admitted cohorts to estimate levels of bed occupancy in the following days. Given the potential of LoS prediction to optimise hospital operations, many methods have been used to predict LoS including machine learning techniques, such as neural networks, logistic regression and regression trees [1] [2] [3], as well as statistical methods including negative binomial regression [4]. Here, we use survival models to obtain the probabilities of discharge of individual patients from daily admissions. Whereas many earlier studies have focussed on a restricted patient group, our aim is a model that can be applied to a wide range of patients.

¹ Corresponding author: School of Electronic Engineering and Computer Science, Mile End Road, London E1 4NS, Email: d.w.r.marsh@qmul.ac.uk

2. Background

The Acute Care Unit (ACU) in the Royal London Hospital provides short term treatment of patients aged 16 or older, admitted non-electively from Emergency Department (ED) or specialist clinics. After 48 hours patients are typically moved to specialist units or discharged. Patients are admitted to the ACU with a wide range of conditions, including injuries, acute health conditions and acute episodes of chronic conditions.

We used the data of the admissions to the ACU in the period 2013-17, consisting of 35,792 admission records for 20,836 unique patients – 52% women and 48% men, with a median age of 65 years. Median LoS was 4 days, with a mean of 8.4 days. 4.8% admissions ended with the death in hospital. There were 6,246 unique ICD-10 diagnostic codes in the sample. Patients' LoS was defined as the time interval between admission to ACU and discharge from hospital or death.

3. Methods

3.1. Data transformation

The Barts Health NHS Trust stores information about patient encounters in a Cerner Data Warehouse. Although the stored data is comprehensive, there are several challenges for building a data set for predictive modelling.

- Multiple identifiers are used: although there is a unique patient ID, every admission to a separate unit, including moves between units during a stay, is recorded with a unique encounter ID. The encounters that make up a hospital stay are not linked: a way to do this was found to assemble a patient's diagnoses, physiological results and treatment data for one admission.
- Data is entered at different times. Some data are entered into the electronic system in real-time during admissions but other data are recorded on paper and only entered into EHR at the end of patient's admission to the hospital. This can result in some of the data not entered into system or data recorded with errors.
- Codes and descriptions of diagnoses are stored together and there is no commonly used system to group individual diagnoses into clusters of clinically similar categories. A patient's diagnoses and physiological tests are often recorded with ED encounter ID, instead of ACU encounter ID.

We extracted demographic data from the ACU unit using patient and admission IDs and cleaned duplicates and errors. We extracted diagnoses data from all hospital's units using the patients ID in the demographic data. For each admission, multiple diagnoses were recorded. To obtain information about comorbidities, we extracted ICD-10 codes from all admissions before the current one, counting as comorbidities diagnoses that were recorded at least once in an earlier admission, excluding any from the latest hospital stay. However, only 33% of admissions were repeats; other patients were assumed free of serious comorbidities. We used the R *icd* package to classify ICD codes into comorbidities and calculated Charlson scores.

To cluster acute diagnoses, we extracted ICD codes and descriptions from current admissions and used R *erm* package, that groups clinically similar diagnoses. For the codes that could not be matched automatically, manual comparisons and corrections were made, from the ICD-10-CM 2014 to the ICD-10-CM-2019 codes, compatible with *erm* package. This method enabled all the common acute diagnoses to be clustered. Physiological tests results (blood tests and NEWS score), were extracted from all

hospital units from time interval *admission date minus one day* and *admission date plus two days*, to account for the tests performed in ED for patients admitted overnight and for repeated tests e.g. when the first sample has been compromised. This procedure resulted in missing values of 3%-5% for the most common tests (e.g. Hemoglobin, RBCDW) and up to 30% for less common ones (e.g. Bilirubin, Albumin). The missing values were imputed with population normal values.

In the last stage, we derived additional variables from the extracted data. For example, for *Kidney_problem* variable, we calculated for repeated admissions median Creatinine for patients' previous admissions and extracted dialyses and kidney transplant codes. We assigned the category 3 for admissions with kidney transplant recorded, 2 for ones with dialyses, 1 with the Creatinine level at a given admission at least 1.5 higher than median Creatinine from previous admissions and 0 if none applied. *Admission_number* and *Readmission_within_year* variables were derived to indicate if an admission was a consecutive one and if a given patient was often readmitted. The final data set consisted of 49 predictors.

3.2. Survival analysis

We developed and evaluated semi-parametric (Cox) and parametric (log-normal) models and Fine and Gray competing outcomes model. We set the evaluation time for 30 days and we defined the 'event' as discharge from hospital. Deaths during admission were right-censored for Cox and parametric models and a competing outcome for the Fine and Gray model. For Cox and parametric models, we fitted splines for continuous variables to account for non-linear relationships with LoS and we used backward variable selection algorithm to select significant predictors (p-value ≤ 0.05 for a global model).

4. Results

The hazards in the global Cox model for the selected variables were not proportional, and stratifying on non-proportional predictors did improve the proportionality result. Full Fine and Grey model did not converge, so this model was developed with the variables selected for Cox model. All models had similar concordance, presented in Table 1.

Table 1 Comparison of models' concordance.

Model	Concordance Index (AUC)
Cox	0.705
Log-normal	0.706
Fine and Gray model	0.703

Given the similar performance, we chose the Cox model for further analysis, as hazard ratios for discharge event could be directly obtained for Cox model regression coefficients, but not from Fine and Gray model [5] [6]. Validation of Cox model with resampling showed there was little overfitting. Predicting discharges by day 4 showed a mean prediction error of 0.033, with underestimation of discharges for the groups with high discharge probability and overestimation of discharges for groups with low discharge probability. Prediction on day 11 showed a slight overestimation of discharges for patients with medium probabilities of discharge, but the mean error decreased to 0.014. Figure 1 shows Cox model calibration performance on days 4 and 11.

4.1. Bed estimation

We calculated means of individual patients’ discharge predictions with confidence intervals (CI) to obtain the number of beds released from daily admissions. Figure 2 shows the predicted output for days 2, 3, 10,15 and 30 from admission of 14 patients. In the external sample of 180 admissions, for days 2 and 3, 66% of observed numbers of beds released were within predicted CI, day 10 - 81%, day 15 - 79% and day 30 - 75%.

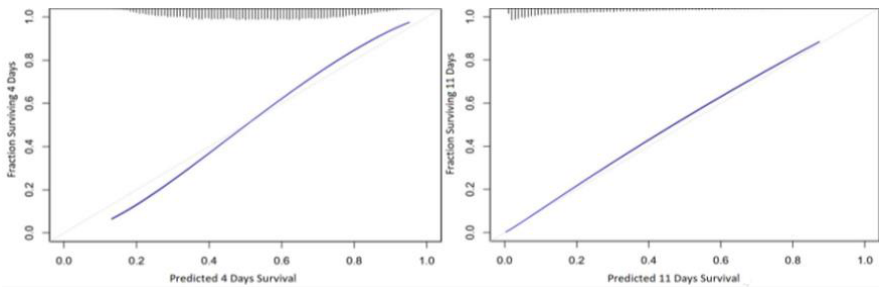


Figure 1 Calibration of predictions on 4th and 11th day - Cox function.

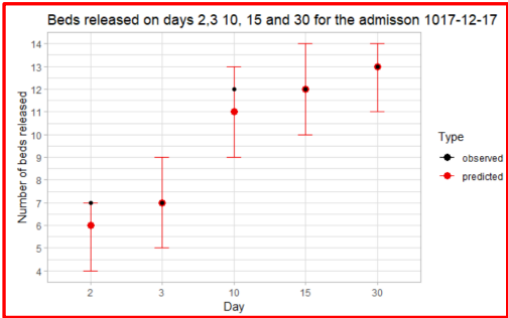


Figure 2 Predicted and observed number of beds released on days 2,3,10,15 and 30.

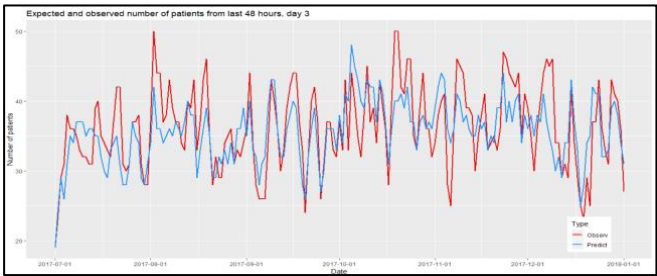


Figure 3 Predicted and observed number of beds occupied from 3 consecutive admissions.

We used the model to predict bed occupancy from several consecutive daily admissions. Figure 3 shows the predicted and observed occupancy after 3 days from the admission on a given day, and the previous 2 days, for 180 admission days. The model predictions match the trends in the observed data. Perhaps because the Cox model does not account for deaths during the admission, the variation in larger in observed than predicted values.

5. Conclusions and future work

We have shown that patient information stored in EHR is sufficient for predicting LoS for a large group of patients with varying diagnoses admitted non-electively. This opens the way to make such predictions routinely without requiring additional data generation.

Building a data sets for modelling is challenging due to the form in which data is stored, but we showed that the necessary linking of records can be achieved so that a uniform dataset can be derived for all patients. We used LoS predictions for patients from daily admissions to estimate the bed occupancy on chosen days post-admission. Despite the suboptimal calibration group of patients with more extreme discharge probabilities, the model performed well for the prediction of beds released from daily admissions.

The next steps in operationalizing the model will be to investigate ways to facilitate real-time entry and access to data. Notably, the ICD-10 codes on which we rely are currently coded manually, but the use of machine learning methods (e.g. [7]) to automate coding has already been investigated and need not be perfectly accurate for LoS prediction. Sourcing comorbidity data for patients admitted to hospital for the first time also needs to be resolved; we plan to investigate the possible use of GP records for this. We will work to extend the prediction to elective admissions and investigate the hospital units to which patients are transferred when they are not discharged from the ACU. We will also investigate ways to incorporate the social factors in LoS analysis, as they can strongly influence a patient's discharge times and cannot be estimated from purely medical records. Finally, to transform the model into a decision tool, senior nurses and bed managers will be consulted on what predictions outputs are useful to support effective bed management.

6. References

- [1] Pei-Fang Tsai, Po-ChiaChen, Yen-YouChen, Hao-YuanSong, Hsiu-MeiLin, Fu-ManLin, Qiou-PiengHuang, "Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network," *Journal of Healthcare Engineering*, 2016.
- [2] Sean Barnes, Eric Hamrock, Matthew Toerper, Sauleh Siddiqui, Scott Levin, "Real-time prediction of inpatient length of stay for discharge prioritization," *Journal of the American Medical Informatics Association*, vol. 23, 2016.
- [3] Lior Turgeman, Jerrold H. May, Roberta Sciulli, "Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission," *Expert Systems with Applications*, vol. 78, 2017.
- [4] Evelene M Carter, Henry WW Potts, "Predicting length of stay from an electronic patient record system: a primary total knee replacement example," *BMC Medical Informatics and Decision Making*, 2014.
- [5] E. Christensen, "Multivariate Survival Analysis Using Cox's Regression Model," *Hepatology*, vol. 7, no. 6, 1987.
- [6] Peter C. Austin, Jason P. Fine, "Practical recommendations for reporting Fine - Gray model analyses for competing risk data," *Statistics in Medicine*, vol. 36, no. 27, 2017.
- [7] YunZhi Chen, HuiJuan Lu, LanJuan Li, "Automatic ICD-10 coding algorithm using an improved longest common subsequence based on semantic similarity," *PLOS One*, vol. 12, no. 3, 2017.

Predicting Postoperative Hospital Stay in Neurosurgery with Recurrent Neural Networks Based on Operative Reports

Gleb DANILOV^{a,1}, Konstantin KOTIK^a, Michael SHIFRIN^a, Uliya STRUNINA^a,
Tatyana PRONKINA^a and Alexander POTAPOV^a

^a National Medical Research Center for Neurosurgery named after N.N. Burdenko,
Moscow, Russian Federation

Abstract. This study aimed to predict the duration of the postoperative in-hospital period in neurosurgery based on unstructured operative reports, natural language processing, and deep learning. The recurrent neuronal network (RNN-GRU) was tuned on the word-embedded reports of primary surgical cases retrieved for the period between 2000 and 2017. A new test dataset obtained for the primary operations performed in 2018-2019 was used to evaluate model performance. The mean absolute error of prediction in the final test was 3.00 days. Our study demonstrated the usability of textual EHRs data for the prediction of postoperative period length in neurosurgery using deep learning.

Keywords. Neurosurgery, Electronic Health Records, Operative Report, Machine Learning, Deep Learning, Recurrent Neuronal Networks, Artificial Intelligence

1. Introduction

Although machine learning is gaining special attraction in neurosurgery during the last years, the text analysis and natural language processing have rarely been used [1], [2]. The operative report is a document describing the details of surgical procedures in a patient's medical records. Written narratively in the majority of clinics, it undoubtedly provides certain knowledge when appropriately explored.

In-hospital stay duration is an indirect estimate of disease severity and healthcare expenses. We hypothesized that the information stored in operative reports could be sufficiently meaningful to predict the duration of the postoperative in-hospital period. To test that hypothesis in a pilot study the predictive model (recurrent neuronal network (RNN) - gated recurrent unit (GRU)) was previously built on 75 531 and tested on 26 123 operative reports [2]. We have shown that the postoperative period could be (to a certain extent) reasonably predicted using the textual description of the surgery, natural language processing, and deep learning. However, our study was limited by the data collected in a period between 2000 and 2017 years [2]. This study aimed to update the

¹ Corresponding author, N.N. Burdenko Neurosurgery Center, Moscow, 4th Tverskaya-Yamskaya str. 16, 125047, Russian Federation; E-mail: glebda@yandex.ru.

model developed previously and validate it on a new set of operative reports never seen by the model before.

2. Methods

In our previous study, the EHRs of N.N. Burdenko Neurosurgery Center was queried to select all operative reports with the corresponded length of in-hospital stay following surgery in the period between 2000 and 2017 [2]. In February 2018 the Hospital information system at N.N. Burdenko Neurosurgery Center was changed. The data on operative reports for a period between January 2018 and June 2019 were obtained from two sources: the old EHR (named “e-Med”; January - February 2018) and the new one (named “Asclepius”; February 2018 – June 2019) [3]. All the incomplete reports and cases with an inaccessible length of in-hospital stay were excluded.

In the current study, all the operative reports dated to 2000-2017 were split into primary surgical cases (the operations performed at our Center for the first time for patients) and repeated surgery, and the training samples were randomly generated from 75% of each subset. The testing samples constituted of the remaining 25% of documents in each category. The model was trained on primary and repeated surgery subsets independently, and its performance was compared in two testing samples accordingly. Finally, the model built on primary cases was fine-tested on operative reports for primary operations dated to January 2018 - June 2019. We did not consider the cases of repeated surgery for the final test due to the reasons explained in section 3.1.

2.1 Text preprocessing

All neurosurgical operative reports typed by neurosurgeons on computer keyboards were retrieved from EHRs. Corpus preparation was done using R programming environment (version 3.5.0) in RStudio IDE for MacOS (version 1.1.453). Raw textual data were preprocessed before fed into deep learning procedure as input:

- Transformed to lower case
- Tokenized with a space separator
- Proceeded by word embeddings

All the tokens generated from the original texts were mapped into numeric vectors from Euclidean space R^d when $d = 400$. These vectors were assumed to capture hidden information about a language, like word analogies or semantic relations between them. Distributed word representation was obtained from operative reports using the FastText library [4], [5]. The idea of that approach was in learning to predict a word from its context or vice versa. We considered the basic vector representations of character trigrams and expanded each word as the sum of the vectors of its trigrams. Then we built the constructions of skip-gram (a set of tokens that imply other tokens between them in an amount not exceeding the preset number) on these sums. Unlike the classic Word2vec and Glove, FastText provides the invariance of embedded tokens to word morphology since N-grams of characters are more common than whole words. As a result, the vector representations we obtained were expected to work efficiently with the rich-in-morphology Russian language.

2.2 The model architecture

The target variable was the postoperative period duration (number of days stored in EHR as integers). To process the operative reports, we RNN-GRU architecture proposed in our previous study [2], [6]. The logarithm of the hyperbolic cosine of the prediction error was used as a loss function since it is resistant to outliers and the second derivative of it exists. The adaptive gradient method RmsProp was utilized as the optimizer with the cut-off gradient value of the maximum norm equal to 1, and the cut-off gradient value outlying the range $[-1, 1]$ [4]. The activation function used was Exponential Linear Unit (ELU) [4]. The hyperparameters of the model were set by the grid search as:

- Batch size = 256 texts
- Dropout (for regularization) = 0.3

The model was trained using a Keras framework (keras.io) on the first 500 vector-represented tokens corresponding to documents. When the number of tokens was < 500 , the sequence was padded with a fixed token complementing sequence to 500.

2.3 The evaluation of model performance

Mean Absolute Error (MAE) was used to measure the quality of model prediction:

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (1)$$

where n – was the number of operative reports, y_+ – true length of the postoperative in-hospital period, y_- – the predicted duration of postoperative in-hospital stay [7].

The model was trained and evaluated within the Python (version 3.6) environment Jupyter Notebook for MacOS.

3. Results

In the period between 2000 and 2017, a total of 77 876 patients (avg. age 39 ± 20 years, males – 55%) underwent 104 506 neurosurgical operations. Initially, we obtained the complete uncorrupted operative reports for 101 664 operations and could verify the duration of postoperative period in 101 654 cases eligible for the inclusion. These texts consisted of 163 316 unique words or 36 438 unique tokens excluding stop-words and words appeared five and fewer times in all reports. The “primary surgery” subset included 75 652 reports (of which 56 739 documents were selected as a training sample). The “repeated surgery” subset included 26 002 reports (with 19 501 used for training).

3.1 Training and validation of the models

The characteristics of samples subsetting for this study and the corresponding measures of deep learning performance are presented in Table 1. The RNN-GRU was applied to “primary surgery” and “repeated surgery” datasets, resulted in MAE of 4.83 and 14.25 days respectively. The model trained on the first dataset outperformed the model previously built on the whole set of data for the period 2000-2017 (Table 1, primary

surgery vs. all reports). However, the result of prediction in the “repeated surgery” dataset appeared to be three times worse. Thus, the RNN-GRU built on the primary surgery dataset was considered as the final model.

3.2 Testing the final model

A total of 2034 operative reports for neurosurgical procedures was obtained from “e-Med” and 11 460 reports were retrieved from “Asclepius” for the period between January 2018 and June 2019. The test of a final model was performed on 11 968 complete operative reports for primary surgery with the known length of the postoperative period (a new testing dataset). The final model demonstrated MAE = 3.00 days (Table 1). The distinct result for “e-Med” data (n = 1 829, MAE = 2.85 days) appeared to be similar for that from “Asclepius” (n = 10 139, MAE = 3.03 days).

Table 1. The results of RNN (bidirectional GRU) testing on operative reports for primary and repeated surgery (n – number of operative reports, AE – absolute error).

	All reports* (2000-2017)	Primary surgery (2000-2017)	Repeated surgery (2000-2017)	Primary surgery** (2018-2019)
Sample size, n	101 654	75 652	26 002	11 968
Range of postoperative stay, days	0 - 1251	0 - 1251	0 - 746	0 - 199
Training sample size, n	75 531	56 739	19 501	11 968
MAE, days	7.78	4.83	14.25	3.00
Predictions with AE = 0 days	14.1%	20.2%	7.7%	20.1%
Predictions with AE ≤ 2 days	54.3%	64.9%	36.3%	67.5%
Predictions with AE ≤ 3 days	63.9%	73.7%	46.8%	78.5%
Median AE, days	2.19	1.52	3.86	1.51
Maximum AE, days	1070.51	731.37	730	174.52

* the results obtained from our previous study [2]

** a new testing dataset containing operative reports for primary surgery performed in Jan 2018 - June 2019

The absolute prediction error in the new testing dataset did not exceed three days in 78.5% of cases.

4. Discussion

To the best of our knowledge, this is the first study in neurosurgery aimed at prediction of the postoperative period based on unstructured medical texts. We chose the operative report for a pilot study to test the utility of textual records. A real-world model production, however, should consider the maximum structured and unstructured data available.

W. Muhlestein et al. (2017, 2018) predicted the length of in-hospital stay after brain tumor surgery using machine learning on non-textual data [8], [9]. Non-elective surgery, preoperative pneumonia, sodium abnormality, or weight loss, and non-White race were found the strong predictors of the prolonged in-hospital stay [8], [9]. In contrast, the results we obtained demonstrate the meaningfulness of narrative medical texts in these non-trivial predictions. The worse predictive capability of repeated surgery reports may indicate that the information essential for prognosis was not captured compared to primary surgery descriptions.

The lower MAE shown in the new testing dataset could be related to the shorter range of the target variable (compared to the development datasets). The high range of AE in all samples was caused by the broad initial range of postoperative periods due to outliers. The outliers in the training samples appeared to be extremely rare and better explained by occasional factors not related to the initial surgery directly, such as severe complications and repeated interventions. The EHR software change was not supposed to affect the results since the operative reports remained in textual format invariant to the system's interface.

The RNN-GRU model was efficient to predict the postoperative period with an acceptable error for the pilot study. However, the rationale for its efficiency should be unraveled in more details, e.g., applying the other machine learning methods, combining it with structured data analysis and considering neurosurgeon's expectation of a postoperative period length. Nevertheless, initial surgical exposure to the human brain appears to be a sophisticated factor itself which might determine the postoperative course and its length. The proposed model may still be improved by text preprocessing (e.g. lemmatization), adding convolutional neural networks layers, optimizing the architecture and training, testing alternative word embeddings (GloVe, Word2vec) [10].

5. Conclusion

Unstructured operative reports served as an informative source for prediction of in-hospital postoperative period in neurosurgery using deep learning. Our study demonstrated the usability of data stored in EHR as texts. *This research was supported by the Russian Foundation for Basic Research (grant 18-29-01052).*

References

- [1] J. Senders, O. Arnaout, A. Karhade, H. Dasenbrock, W. Gormley, M. Broekman, T. Smith, Natural and Artificial Intelligence in Neurosurgery: A Systematic Review, *Neurosurgery*, **83**, (2018), 181–192.
- [2] G. Danilov, K. Kotik, M. Shifrin, U. Strunina, T. Pronkina, A. Potapov, Prediction of Postoperative Hospital Stay with Deep Learning Based on 101 654 Operative Reports in Neurosurgery, *Stud. Health Technol. Inform.* **258** (2019), 125–129.
- [3] M. A. Shifrin, E. E. Kalinina, E. D. Kalinin, A sustainability view on the EPR system of N.N. Burdenko Neurosurgical Institute, *Stud. Health Technol. Inform.* **129** (2007), 1214–1216.
- [4] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, Bag of tricks for efficient text classification, *arXiv Prepr. arXiv1607.01759*, 2016.
- [5] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, T. Mikolov, Fasttext. zip: Compressing text classification models, *arXiv Prepr. arXiv1612.03651*, 2016.
- [6] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, *arXiv Prepr. arXiv1412.3555*, 2014.
- [7] J. Fürnkranz, Mean Absolute Error, *Encyclopedia of Machine Learning*, Springer US (2011), 652–652.
- [8] W. E. Muhlestein, D. S. Akagi, J. M. Davies, L. B. Chambless, Predicting Inpatient Length of Stay After Brain Tumor Surgery: Developing Machine Learning Ensembles to Improve Predictive Performance, *Neurosurgery*, **85** (2019), 384–393.
- [9] W. E. Muhlestein, D. S. Akagi, S. Chotai, L. B. Chambless, The Impact of Race on Discharge Disposition and Length of Hospitalization After Craniotomy for Brain Tumor, *World Neurosurg.*, **104** (2017), 24–38.
- [10] G. Danilov, M. Shifrin, U. Strunina, T. Pronkina, A. Potapov, An Information Extraction Algorithm for Detecting Adverse Events in Neurosurgery Using Documents Written in a Natural Rich-in- Morphology Language, *Stud. Health Technol. Inform.*, **262** (2019), 194–197.

Probabilistic Approaches to Overcome Content Heterogeneity in Data Integration: A Study Case in Systemic Lupus Erythematosus

Alexia SAMPRI^a, Nophar GEIFMAN^a, Helen LE SUEUR^a, Patrick DOHERTY^b,
Philip COUCH^a, Ian BRUCE^{b,c} and Niels PEEK^{a,c},

on behalf of the MASTERplans Consortium

^a*Division of Informatics, Imaging and Data Sciences University of Manchester,
Manchester, UK*

^b*Division of Musculoskeletal & Dermatological Sciences, University of Manchester,
Manchester, UK*

^c*NIHR Manchester Biomedical Research Centre, University of Manchester,
Manchester Academic Health Science Centre, Manchester, UK*

Abstract. Integrating data from different sources into homogeneous dataset increases the opportunities to study human health. However, disparate data collections are often heterogeneous, which complicates their integration. In this paper, we focus on the issue of content heterogeneity in data integration. Traditional approaches for resolving content heterogeneity map all source datasets to a common data model that includes only shared data items, and thus omit all items that vary between datasets. Based on an example of three datasets in Systemic Lupus Erythematosus, we describe and experimentally evaluate a probabilistic data integration approach which propagates the uncertainty resulting from content heterogeneity into statistical inference, avoiding the need to map to a common data model.

Keywords. Probabilistic data integration, content heterogeneity, missing data, biomedical data harmonisation

1. Introduction

Integrating data from different sources into a homogeneous data resource creates powerful opportunities to study human health. However, disparate data collections are often heterogeneous, which complicates their integration.

Systemic Lupus Erythematosus (SLE) is a chronic autoimmune disease of unknown aetiology, affecting approximately 16,000 people in the UK. Through the MASTERplans programme [1] we have gained access to data from three cohort studies: Aspreva Lupus Management (ALMS), Lunar, and Exploratory Phase II/III SLE Evaluation of Rituximab (Explorer).

Content heterogeneity refers to the situation in which data items that are not equally represented across different data sources [2]. In the MASTERplans example, content heterogeneity occurs because Lunar and Explorer did not record age at onset of disease,

smoking status, white blood cells (WBC), platelets and lymphocytes, creating missing variables problems. Another form of content heterogeneity occurs because ethnicity was not recorded using the same categories across the three studies.

Traditionally, people solve these problems by mapping all datasets to a common data model. This model would only include variables that are present in all datasets, and, in case of granularity problems, categories that exist in all datasets.

2. Data and Method

First, we select all the datasets that are available and can enable us to answer the research question based on the description in the metadata. For the datasets that do not satisfy all the selection criteria concerning content heterogeneity we build predictive models that estimate the probability that these criteria are present based on other information that is available, using more complete datasets. The probabilistic relationship selection criteria are learned from other datasets. Our approach aims to preserve all the available information in the original datasets by translating content heterogeneity problems into missing value problems, and then solving these missing value problems using established methods (multiple imputation). The method assumes that naming differences are resolved beforehand, that the original datasets were sampled from the same population and that data are missing at random.

We applied our method to the three SLE datasets, to answer the research questions “does ethnicity predict response to treatment?” and “Finding the set of variables that best predict drug response”. For the first question a granularity problem occurred, and for the second question a missing variables problem occurred. For both solutions, we applied imputation using multivariate imputation by chained equations (MICE) [3] and a random forest approach called missForest [4].

In order to answer both questions, we included only ALMS-M (maintenance) information from the ALMS study data since we required patients to have a 12-month follow-up visit (with disease severity evaluated) which was not available in the ALMS-I set (which followed patients for only 6 months). We rescaled the first visit as zero days and calculated days for each following visit relative to that. For both questions were used datasets that included patients from all the three studies together, with multiple visits per subject. For the response measure only, in order to have only one visit per patient, we kept the visit that had the least absolute difference from 365 days (12 month). First question’s dataset consists of the 545 patients and 9 variables i.e., gender, age, ethnicity, height, weight, drug response (BILAGScore total), creatine, body mass index (BMI), and current treatment. Content heterogeneity occurs because ethnicities’ levels are not the same across all the sources. Ethnicity in ALMS has 4 levels, i.e., ‘Caucasian’, ‘Asian’, ‘Black’, and ‘Other’, where ‘Other’ are also different levels of ethnicity. On the contrary, ethnicity’s levels in Lunar and Explorer are ‘Caucasian’ and ‘Black or African American’. Therefore, in ALMS, ethnicity’s granularity is very high. For the granularity problem we add a categorical variable that classifies patients based on the ‘Other’ ethnicity. With this additional variable we eliminate misclassification.

The dataset with the missing variables problem contains not only the initial variables mentioned earlier but also smoking status, lymphocytes, WBC, and platelets. Missing variable problem occurs because these extra variables were included only in Lunar and Explorer data. We remove records for 7 patients with missing value for height and 9 with

missing creatine leading to 530 patients. This way makes it better and clearer to understand how the method works in application.

The chosen parameters for missForest imputation are 100 trees and a maximum of 50 iterations for both research questions. For MICE the chosen parameters are 30 and 20 imputed datasets, 50 and 40 iterations, and 10,000 and 15,000 seed for the research questions respectively. At the end, we fit linear regression models to the complete datasets that resulted from both imputation methods.

3. Results

In Table 1, we see the coefficients for the linear regression model that estimates the drug response based on ethnicity. We compare the coefficients from both imputation methods; MICE and random forest. Table 2 shows the coefficients for the linear regression model that estimates the drug response based on ethnicity, gender, smoking status, BMI, treatment, creatine, lymphocytes and platelets. Figure 1. shows the distributions

Table 1. Coefficients for linear regression model in granularity problem

	estimate		std error		pvalue	
	MICE	MissForest	MICE	MissForest	MICE	MissForest
Intercept	1.1917	1.0000	4.9643	6.1870	0.8104	0.8720
Asian	3.8140	4.6710	5.0076	6.2090	0.4467	0.4520
Black or African American	5.8783	6.0200	5.0009	6.2180	0.2404	0.3330
Cape Coloured	-0.9833	-1.0000	7.5161	8.7500	0.8960	0.9090
Caucasian	6.9757	7.2210	4.9789	6.1990	0.1618	0.2450
East Indian	10.3623	5.0000	5.4224	8.7500	0.0567	0.5680
Eritrean	12.4797	11.0000	5.7933	8.7500	0.0319	0.2090
Hispanic	4.1158	4.5000	6.2853	7.5780	0.5129	0.5530
Mexican Mestizo	2.8448	3.2610	5.0588	6.3200	0.5741	0.6060
Middle Eastern	-0.8167	-1.0000	7.5825	8.7500	0.9143	0.9090
Mixed	2.7500	1.0000	6.5575	8.7500	0.6751	0.9090
Moroccan	5.9950	5.0000	7.4419	8.7500	0.4210	0.5680
Native American	4.0167	2.0000	6.9421	8.7500	0.5632	0.8190
Nicaraguan	9.7708	6.0000	6.7114	7.1440	0.1462	0.4010

Table 2. Coefficients for linear regression model in missing variables problem

	estimate		std error		pvalue	
	MICE	MissForest	MICE	MissForest	MICE	MissForest
(Intercept)	5.4982	6.2212	1.89E+00	1.9747	5.76E-03	0.0017
Caucasian	1.2472	1.2619	7.47E-01	0.7485	1.22E-01	0.0924
Other	-0.5385	-0.5426	8.00E-01	0.7867	4.37E-01	0.4907
Male	-2.0917	-1.9727	8.64E-01	0.8238	1.89E-02	0.017
Never smoker	-0.4917	-0.2473	8.73E-01	0.8431	5.57E-01	0.7694
Previous smoker	-1.0473	-1.0190	1.02E+00	0.9793	3.10E-01	0.2986
BMI	0.0402	0.0465	4.33E-02	0.0421	4.17E-01	0.2692
Placebo + MMF OR MMF	-0.2093	0.1268	7.95E-01	0.6919	8.74E-01	0.8547
Placebo + MTX OR MTX	3.4207	3.8444	1.54E+00	1.5080	2.73E-02	0.0111
RITUX + AZA	1.4513	1.5648	1.18E+00	1.1227	2.00E-01	0.1634
RITUX + MMF	0.6037	0.8455	9.02E-01	0.7961	5.08E-01	0.2888
RITUX + MTX	2.0002	2.2410	1.24E+00	1.1906	1.20E-01	0.0604
Creatine	1.0438	1.0278	9.24E-01	0.8998	3.23E-01	0.2539
Lymphocytes	-2.3277	-2.9186	5.60E-01	0.5525	2.40E-05	1.89E-07
Platelets	0.0090	0.0068	3.40E-03	0.0036	1.11E-02	0.05723

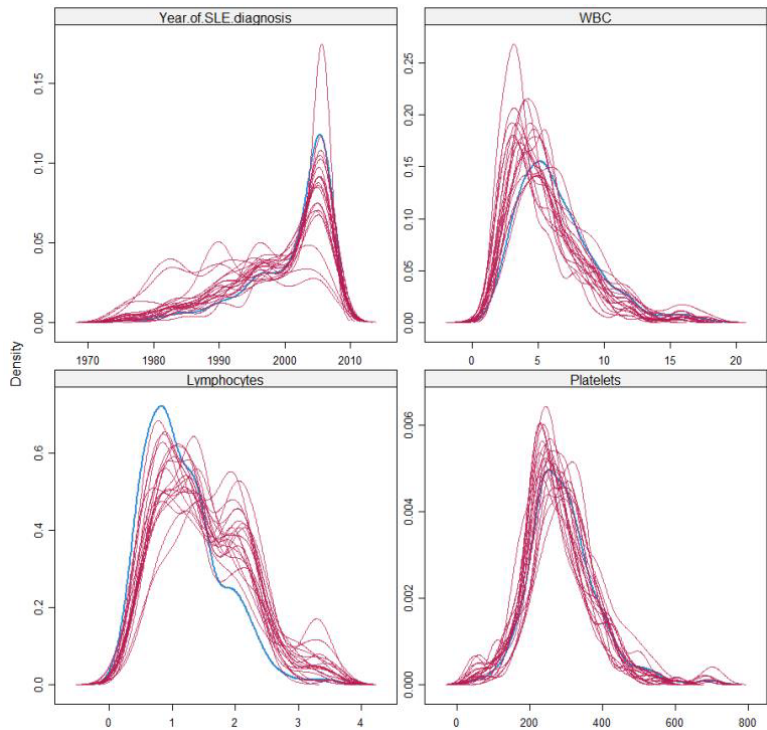


Figure 1. Density plots for year of SLE diagnosis, WBC, lymphocytes and platelets, in missing variables problem. Blue line shows the observed data and the red lines the imputed data from each of the imputations in MICE.

4. Discussion

The impact of ethnicity on drug relationship has the huge potential of achieving and improving the efficacy of results in precision medicine and differences in recommended drug doses [5, 6]. In table 1 we see that ‘Eritrean’ is significant and it implies that ethnicity is associated with drug response. Concerning the imputation methods, MICE did not have any misclassifications. However, missForest misclassified 1 patient as ‘Caucasian’ and 2 as ‘Black of African American’. Further analysis is suggested such as application of our method to other data sources.

In Table 2, the coefficients of linear regression models, after imputation using MICE, show that all the variables included in the model are significant. Nonetheless, imputed values with missForest imputation show less significance. Gender, treatment and lymphocytes are the most significant with missForest imputation. Figure 1 also shows the distributions of the observed and imputed values for missing variables problem. As we see in figure 1 extreme values affect the shape of the plots. However, the central tendencies of the density plots of imputed data appear relatively similar to the observed ones.

In general, our approach contains a number of important ingredients. First, it insists on the future existence of health data heterogeneity. Therefore, our approach strives for

post alignment rather than pre-alignment of Big-health/bio datasets. Second, as a post-alignment of heterogeneous data sources will be always imperfect and it is not a problem that datasets are not content equivalent, if they estimate the probability that they are. Third, this approach is pragmatic in the sense that always provides an answer- although it might not better when the source datasets do not provide useful information to answer the research question. However, the results of the probabilistic data integration would be the same as those that would result from analysing an integrated dataset.

5. Conclusion

Existing methods for dataset integration rely on mapping to common data models, often resulting in a substantial loss of information that is present in the source datasets. One promising alternative relies on probabilistic methodologies. This paper has illustrated this approach using a real-world example from Lupus cohort studies. Rather than relying on perfectly harmonised data items, our method propagates the uncertainty that results from imperfect harmonisation into the statistical analysis, thus obviating the need for data integration through a common data model.

Ideally, shared data models would be implemented at source, enabling uniform data collection at different sites and studies. But in reality, data standardisation is always imperfect, and our approach embraces this imperfection rather than trying to extinguish it. Future work includes expanding of the general applicability of the method, and comparing results of the proposed integration techniques with gold standard results through statistical simulation studies. Moreover, to demonstrate its utility the developed approach will be applied to real world biomedical and health datasets such as the studies in SLE that were used for our examples.

Acknowledgement

This study received financial support from the Engineering Physical Sciences Research Council (EPSRC) Doctoral Training Partnerships.

References

- [1] MASTERPLANS Maximising SLE Therapeutic Potential by Application of Novel and Systematic Approaches [Internet]. Lupusmasterplans.org. 2019 [cited 14 October 2019]. Available from: <http://www.lupusmasterplans.org/home.html>
- [2] W. Sujansky, Heterogeneous Database Integration in Biomedicine, *Journal of Biomedical Informatics*, 34 (2001), 285–298.
- [3] S. van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate Imputation by Chained Equations in R, *Journal of Statistical Software* 45 (2011).
- [4] MissForest—non-parametric missing value imputation for mixed-type data, *Bioinformatics* 28 (2011), 112–118.
- [5] S.U. Yasuda, L. Zhang, S.M. Huang, The Role of Ethnicity in Variability in Response to Drugs: Focus on Clinical Pharmacology Studies, *Clinical Pharmacology* 84 (2008), 417–423.
- [6] R.R. Shah, A. Gaedigk, Precision medicine: does ethnicity information complement genotype-based prescribing decisions? , *Therapeutic Advances in Drug Safety* 9 (2018), 45–62.

Problems in FAIRifying Medical Datasets

Matthias LÖBE^{a,1}, Franz MATTHIES^a, Sebastian Stäubert^a, Frank A. Meineke^a, and Alfred WINTER^a

^a*Institute for Medical Informatics, Statistics and Epidemiology (IMISE), Leipzig University, Leipzig, Saxony, Germany*

Abstract. Despite their young age, the FAIR principles are recognised as important guidelines for research data management. Their generic design, however, leaves much room for interpretation in domain-specific application. Based on practical experience in the operation of a data repository, this article addresses problems in FAIR provisioning of medical data for research purposes in the use case of the Leipzig Health Atlas project and shows necessary future developments.

Keywords. FAIR data, Research data management, Data sharing

1. Introduction

The FAIR Guiding Principles for scientific data management have been widely accepted since their publication in 2016 [1]. Nevertheless, their practical implementation is difficult, especially in the area of particularly sensitive personal medical data. While many data repositories and data management solutions claim to be FAIR, the qualitative implementation of the FAIR principles varies [2]. A typical university hospital supports various types of research projects in addition to patient care, such as clinical trials for the approval of new drugs or therapy optimization, disease-specific registries for the most complete possible registration of all patients suffering from a specific disease, population-based cohorts for epidemiological long-term monitoring of the population, biomaterial banks, molecular genetic analysis (OMICS) and much more. Most of these research projects generate high quality data in terms of completeness, consistency and accuracy. Data is collected, curated, analyzed; but rarely used again. The objective of this contribution is to show the problems that arise when building a research data repository in accordance with the FAIR principles, with special emphasis on the health research domain.

2. Method

Medical research on humans is resource-intensive and ethically challenging, so it should be made possible to get the maximum out of the data. The Leipzig Health Atlas is a research data repository that has been built up over the last few years [3] and is now constantly being filled with content. Based on the SEEK platform [4], not only

¹Corresponding Author: Matthias Löbe, Härtelstraße 16-18, 04107 Leipzig, Germany; E-mail: matthias.loebe@imise.uni-leipzig.de

data sets but also other types of research artefacts such as systems biology pipelines, bioinformatics models or medical ontologies will be managed. In addition to the core task of long-term archiving data are to be made available to other researchers ("Data Sharing") and analyses of data should be repeatable by third parties ("Reproducibility"). The method empirical-analytical approach used here is based on our experience in operating the portal and the consulting and feedback of scientists when uploading their results.

3. Results

This chapter outlines various FAIR hurdles for biomedical research, broken down according to the main axes of the FAIR principles, which have arisen several times and for which proposals are made for dismantling them.

3.1. Considerations for FAIR axis 1: Findability

Findability deals with the problem of recognising the existence of resources and referencing them. Many data repository software systems or content management systems assign (locally) unique numbers by themselves, which could be made globally unique by a suitable prefix. However, the required "eternal persistence" cannot be guaranteed with regard to updates or system changes. This results in the need for special digital identifier registry software. Such systems already exist in the medical field, e.g. the HL7 Object Identifier (OID) Registry as defined in the recommendation ITU-T X.660 and ISO/IEC 9834 series. Unfortunately, the OID system [5] is only one of many possible handle systems, Digital Object Identifiers (DOI) are widely used in scientific journals, and Uniform Resource Identifiers (URI) are used in distributed web-based software systems. While each of these systems can be designed to be *resolvable* in principle, there is no agreement on other desirable characteristics [6] such as descriptive human readable designation as part of identifiers.

Another problem is the granularity at which data is to be assigned its own identifiers. It is undoubtedly important, for example, to provide a data set that was the basis of a scientific publication in one version with its own identifier. However, medical databases are characterized by frequent and extensive changes, whereby data is not only added, but also updated. A continuous versioning would be impracticable here. A further problem is that it is desirable with regard to the use of data by external systems to also be able to reference only parts of data sets or single data elements. In the sense of the idea of a Semantic Web, each information atom would get its own URI, which is, however, insufficiently supported in practice by existing software.

While the actual search can be very well supported by existing software using keywords, ontologies, or categorical filters, there is a lack of community accepted metadata vocabularies for describing data sets (such as MIAME – Minimum information about a microarray experiment or the WHO Clinical Trial Registration Data Set) for many areas of biomedical research, which often makes it difficult for users to navigate within a repository. Furthermore, due to the growing number of repositories, it would be desirable if these were federatively linked so that a user does not have to formulate a request several times in different places and aggregate the results. Here, too, there is a lack of a unifying data catalog specification for available

health data describing the content of data collections and the location. Ultimately, individual data elements must be described very precisely in order to prevent misinterpretation. Simple designations cannot do this. Metadata repositories (MDR) allow semantic specification of clinical concepts through annotation with medical terminologies, as well as the precise specification of units of measurement, formats, data types, and reference ranges. Different systems exist, but their use is very complex and does not fit into the regular work processes, which limits their application [7].

3.2. Considerations for FAIR axis 2: Accessibility

Accessibility enables a researcher to actually retrieve data. To have metadata permanent available, even when the data are no longer available, is quite easy to achieve technically but harder to enforce culturally because some researchers have not yet fully embraced the meaning of this criterion. Access via a standardized communication protocol is also not a technical problem since Representational State Transfer (REST) interfaces based on the HTTP protocol have established themselves and current open standards such as HL7 Fast Healthcare Interoperability Resources (FHIR) [8] are based on them. The difficulties with access to personal medical data lie rather in their sensitive nature for the individual patient. Comprehensive regulations such as the European General Data Protection Regulation (GDPR) regulate in detail the conditions and restrictions for their collection and further processing. In general, the use of such data for research purposes requires the informed consent of the patient or proband. The declaration of consent must be sufficiently specific with regard to the research objective, the persons accessing the data and the circumstances of the data processing, which can prevent subsequent data sharing. Also, identifying data must not be stored together with other research data, which makes subsequent data collection or record linkage more difficult. Furthermore, participation in research projects is always voluntary and can be recalled at any time. Anonymisation procedures for individual medical data can remove these restrictions, but are accompanied by a great loss of information entropy. For this reason, it is necessary to create and use further harmonised metadata vocabularies on topics such as the legal basis for data collection or the different variants of Informed Consent.

Since such data will rarely be automatically accessible, the process of requesting data access and secure data retrieval plays an important role. In practice, a simple reference to a contact person for each data record represents a major hurdle for both the data holder and the interested party, since medical researchers are not always fully aware of the legal framework. The establishment of a Data Access Board for checking and approving data use proposals according to a usage concept is strongly recommended. The data repository should provide an authentication service for internal and external users based on a rights and roles concept. To address the concerns of data owners about possible breaches of data privacy during data sharing, contact points for questions regarding data sharing, data protection, etc. should be established.

Further problems regarding permanent accessibility are contradictory rules for the storage of data. While individual funders consider a storage period of at least 10 years as part of good scientific practice to be appropriate, some data usage agreements state that the shared data must be deleted immediately after the end of the project, which has already led to data records no longer being available shortly after the publication of a scientific article. A regulatory harmonisation must take place here.

3.3. Considerations for FAIR axis 3: Interoperability

Interoperability is the ability of distributed, heterogeneous systems to exchange information in an unambiguous and exploitable manner. There are countless standards, conventions and best practices in biomedical research, but in many cases researchers take advantage of the freedom of science to act according to their own ideas. For the operators of research data repositories, this often means that only minimal ideas about data structures can be enforced (e.g. a tabular format with character-separated values).

Since technical and semantic interoperability are generally subject to higher requirements, operators should support data owners in transforming the data into a widely applicable format for knowledge representation. Such a FAIRification process requires the definition of a rich target data model. Various common data models are suitable for this; the selection can be influenced by later intentions of use. The latest, but also most universally applicable model is the already mentioned HL7 FHIR. It covers practically all areas of medicine, even if some resources are not yet finally specified. In the area of clinical trials, the standards of the CDISC organization are of importance, especially the Operational Data Model (ODM) [9] in conjunction with the Study Data Tabulation Model (SDTM). For large, distributed data analyses, the OHDSI Observational Medical Outcomes Partnership (OMOP) [10] model has gained great popularity. For all these models, however, rudimentary ETL tools are available at best.

International medical terminologies such as ICD-10, LOINC or SNOMED CT are very suitable for describing clinical concepts in detail. They only partially meet the requirement that vocabularies should also comply with FAIR principles. Terminology services can aid the linking process.

3.4. Considerations for FAIR axis 4: Reusability

Reusability describes the application of data for secondary purposes independent of the original use. With regards to medical data, provenance is the most important topic of the FAIR reusability metric. While provenance is a broad topic and the demarcation to data acquisition is not sharp, data owners should specify in detail the circumstances of the data collection and processing, i.e. data sources, data validation rules, format conversions, data cleansing, derived or aggregated data, measuring instruments, scripts, software libraries, observers. This greatly increases the confidence of external researchers in the data sets. A further recommendation is the provision of simple web-based visual analysis tools such as tranSMART [11], which give potential interested parties an overview of the depth of the available data and so enhances reusability.

3.5. Additional considerations for medical FAIR data

Some aspect does not fit into one of the FAIR data categories but are nevertheless part of our recommendations: 1) Data quality is a primary concern when one relies on external data. All procedures to ensure high data quality (technical validations, manual curation) should be made explicit. 2) If datasets are not eligible for sharing (privacy, volume), privacy-preserving data analysis techniques could be an option. 3) Repository operators should provide additional services to facilitate sharing and usage of data e.g. pseudonymization, de-identification, anonymization, record linkage. 4) The effort and benefit of FAIR verification of data is generally unevenly distributed at the expense of

the data owner. However, an incentive for them may be to receive feedback on improved or additionally calculated data and to add it to the original data source.

4. Discussion and conclusion

The FAIR data principles can well be applied to medical data repositories, yet some restrictions will apply, mostly with regard to accessibility/data privacy. The LHA project is work in progress; not all of our recommendations are solved, implemented and evaluated yet. Even through some of our findings lack a broad, independent evaluation, we expect most of them to be translatable to similar projects. A first draft of a guideline for implementing a general FAIR open data policy in health research was developed as part of the FAIR4Health project [12].

The greatest difficulties for FAIRification currently lie in the lack of availability of community-consented vocabularies and powerful tools for transforming data to common data models like HL7 FHIR. While several initiatives (e.g. FORCE11, GOFAIR, Research Data Alliance) assess the coverage of the FAIR principles of different systems and publish FAIR metrics onto how these principles should be put into practice, effort and costs are currently too high for many common data collection projects. Incentives are needed to take on the challenge of FAIRification. Furthermore, there should be more standardization on processes for data access and data extraction.

Acknowledgements: This work was supported by the European Union (H2020/824666 – FAIR4Health), the German Ministry of Education and Research (LHA: 031L0026, SMITH: 01ZZ1803A) and the German Research Foundation (NMDR: WI 1605/10-1).

References

- [1] M.D. Wilkinson, M. Dumontier, I.J.J. Aalbersberg et.al., The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3 (2016), 160018.
- [2] A. Dunning, M. de Smaele, and J. Böhrer, Are the FAIR Data Principles fair? *IJDC 12* (2017), 177–195.
- [3] F.A. Meineke, M. Löbe, and S. Stäubert, Introducing Technical Aspects of Research Data Management in the Leipzig Health Atlas. *Stud Health Technol Inform* 247 (2018), 426–430.
- [4] K. Wolstencroft, S. Owen, O. Krebs et.al., SEEK: a systems biology data and model management platform. *BMC Syst Biol* 9 (2015), 33.
- [5] S.J. Steindel, OIDs: how can I express you? Let me count the ways. *J Am Med Inform Assoc* 17 (2010), 144–147.
- [6] J.A. McMurtry, N. Juty, N. Blomberg et.al., Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biol* 15 (2017).
- [7] M. Löbe, User Expectations of Metadata Repositories for Clinical Research. *Stud Health Technol Inform* 253 (2018), 60–64.
- [8] M.L. Braunstein, Healthcare in the Age of Interoperability: The Promise of Fast Healthcare Interoperability Resources. *IEEE Pulse* 9 (2018), 24–27.
- [9] V. Huser, C. Sastry, M. Breymaier et.al., Standardizing data exchange for clinical research protocols and case report forms: An assessment of the suitability of the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM). *J Biomed Inform* 57 (2015), 88–99.
- [10] G. Hripcsak, J. Duke, N. Shah et. al., Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform* 216 (2015), 574–578.
- [11] B.D. Athey, M. Braxenthaler, M. Haas et.al., tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. *AMIA Jt Summits Transl Sci Proc* 2013 (2013), 6–8.
- [12] T. Hernández-Pérez, E. Méndez Rodríguez: D2.3. Guidelines for implementing FAIR Open Data policy in health research. Available at: <https://www.fair4health.eu/en/resources/project-deliverable>

Prototyping a Tool for Processing Genetic Meta-Data in Microbiological Laboratories

Jan AESCHIMANN^a, Silvan HUBER^a, Daniel WÜTHRICH^{b,c}, Helena SETH-SMITH^{b,c}, Jürgen HOLM^a, Thomas BÜRKLE^a, and Murat SARIYAR^{a,1}

^aBern University of Appl. Sciences, Dept. Medical Informatics, Switzerland

^bClinical Bacteriology and Mycology, University Hospital Basel, Switzerland

^cApplied Microbiology Research, Department of Biomedicine, University of Basel, Switzerland

Abstract. Next generation sequencing (NGS) technologies allow improved understanding of pathogens. In the upstream processing of generating genomic data, there is still a lack of process-oriented tools for managing corresponding meta data. In this paper, we provide a description of how a process-oriented software prototype was developed that allowed the capture and collation of metadata involved when doing NGS. Our question was: How to develop an interactive web application that supports the process-oriented management of genetic data independent of any sequencing technique?

Keywords. data management; microbiological data; ngs; web application

1. Introduction

Modern microbiological laboratories use next generation sequencing (NGS) for understanding the genomes of pathogens. This allows, for example, the detection of outbreaks and the determination of antibiotic resistance mechanisms. In order to generate annotated genomic data, bioinformatic and biomedical knowledge are necessary. For NGS analysis, software solutions such as Ridom Seqsphere+ [1], CLC Genomics Workbench [2] or UGENE [3] can be used and facilitate the analysis of sequence data, annotation, alignments, and visualization such as phylogenetic trees. However, there is more information involved in the upstream processing of generating genomic data, which are often not fully considered in such approaches [4].

The collection and reliable storage of metadata relating to the sample is critical, particularly in an ISO accredited clinical microbiology laboratory, but also of relevance to other laboratories performing NGS. Three steps for generating a genome-based laboratory typing report from a clinical sample can be distinguished: sample preparation, sequencing and analysis. In the sample preparation step, clinical samples or isolates are subject to DNA extraction. In this stage, relevant data relating to the patient, the sample, and the extraction are collected. In the sequencing step, data on the concentration of DNA, the library method, the sequencing platform used, and a quality

¹ Corresponding Author, Murat Sariyar, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: murat.sariyar@bfh.ch.

control of the resulting sequences are annotated. In the analysis step, the sequences are interpreted, for which biomedical information is indispensable.

The workflow of capturing and collating metadata during NGS was analyzed in the Division of Bacteriology at the University Hospital Basel. The processes for capturing the data associated with NGS were examined before starting the process of software development. At baseline, Excel worksheets on a server were used with no established workflow of how and when to commit changes or new entries. Our central goals were to facilitate NGS metadata management in order to reduce the likelihood of errors in the database and allow for parallel editing of different parts of the database. The guiding research questions was: How to develop an interactive web application that allows an efficient process-oriented management of genetic data independent of any sequencing technique?

2. Methods

With a Pubmed and Google Scholar research using the keywords “metadata microbiology”, “ngs metadata”, “pathogen ngs metadata”, “metadata management system” and “metadata management system genome” we checked if an external tool for our goals is available. Next, we iteratively created a mockup together with the laboratory staff using the Figma mockup tool [5] in order to define an optimal workflow. The subsequent requirement engineering process was based on IEEE 830-1998 [6] and lead to the definition of the software architecture.

In the development phase, we used the agile SCRUM technique [7] with seven sprints lasting one week each. We made use of the JavaScript node.js (loopback) framework for the backend and the Vue.js framework for the frontend. The REST-based philosophy of many JavaScript frameworks is not designed for multiuser systems that require isolated updates and deletion of common resources. Therefore, we used *socket.io* as a JavaScript library for real-time web applications that enables bi-directional communication between web clients and servers, ensuring that locks on datasets can be propagated beyond one session within a REST-based access [8]. For usability tests based on a structured questionnaire and for final modifications according to the answers, two more weeks were scheduled.

3. Results

In our analysis of the management of genetic meta data, we identified six central process steps concerning the weekly sample preparation workflow. Initially, a laboratory order from internal or external customers arrives together with the material to be sequenced. The order is validated and prioritized. In step 2, the sample type (isolate or DNA) is checked. Step 3 selects the necessary DNA extraction process with respect to the material, if required. Step 4 selects the number of samples (4, 24, 48, or 96 depending on the sequencing device and cartridge). After performing the DNA extraction, the sequencing can be planned with the information generated during extraction (step 5). Finally, the samples for the next NGS run are selected, a sequencing run is performed, and sequencing data is saved for later analysis and reporting (step 6).

These process steps are divided into four different views according to the data input required. For steps 1-2, it is the view “*Planned*” with data on the requesting facility, the patient, the type of the material, the date of receiving the order, the type of the pathogen, the priority of the order (from “A”: high to “D”: low); the view for step 3 is “*extracted*” with the same data as in the former view and DNA concentration; the view for steps 4 and 5 is “*Preparing run*” with data on the finalized DNA extracts to be sequenced; for the last step the view is “*Sequenced*” with data on the sequencing device used, date of sequencing and some of the sequencing properties, e.g. adaptor oligonucleotides ligated to fragments of the DNA for forming the DNA library.

During the requirements analysis, we also identified the original data management workflow based on Excel worksheets, which showed several disadvantages:

- Lack of information on the order of inserting data into the Excel worksheets
- No import options for pre- formulated orders, potentially leading to errors in manual entries
- No export options for selected data, which could lead to errors in manual copy/paste operations during report generation
- No support for batch processing of the data
- No possibility of representing a complete repetition of a sequence run by shifting metadata from the view “*sequenced*” to former views

To implement a web application that addresses these disadvantages and provides an affirmative answer to our research questions (see introduction), we considered additional functional and non-functional requirements, e.g. a table view for general overview or ubiquitous access to the system from everywhere within the laboratory, which mandates a web interface that is compatible with current browsers. The main non-functional requirement was user guidance through the metadata management process by prompting only for relevant fields in each step and by providing help functions, sorting options, search functions as well as tooltips. The prototype is called “Red Maple”, and the start page is given in Figure 1. The UML sequence diagram for our solution using web sockets is presented in Figure 2.



Figure 1. Start page of the prototype *Red Maple* with the four views “Planned”, “Extracted”, “Run”, “Sequenced” and a summary of datasets included in them. The bubbles in the right column represent the links for details within these views.

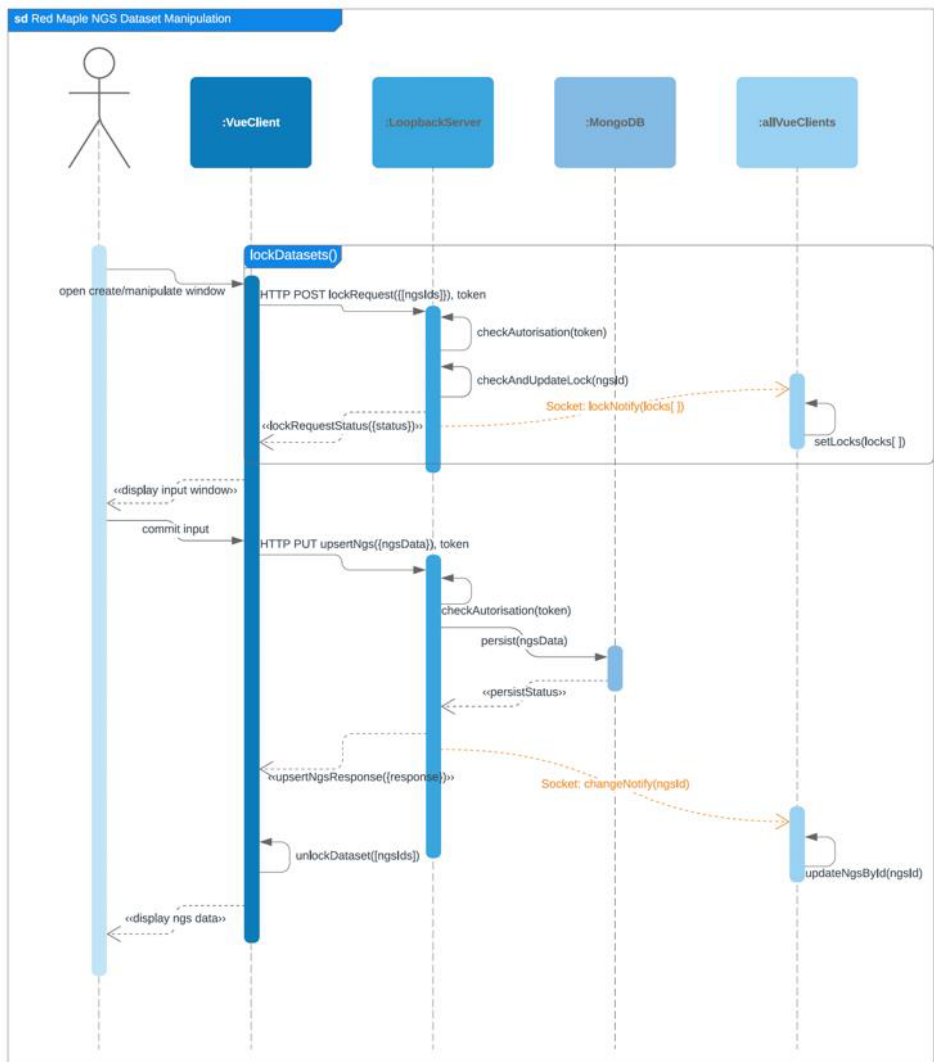


Figure 2. The UML sequence diagram shows how data is accessed via the frontend (VueClient). In case of manipulation of the data, tokens are used, which allow manipulation of data in the backend (MongoDB is the NoSQL database used by the loopback Server). Locks are necessary in order prevent anomalies possible when the same data is tried to be manipulated simultaneously.

4. Discussion

Typically, laboratory analyses are associated with clear diagnostic procedure steps described in medical guidelines. The NGS processes within microbiological laboratories are different in that they are often not directly based on a medical guideline but are necessary to gain information on transmission and outbreaks, thus affecting

more than one single patient. This is one of the reasons why such NGS analysis processes are difficult to integrate into the workflow of existing laboratory information systems. Hence, instead of stretching existing LIS solutions, it seems desirable to develop a modular solution that is useful for many microbiological laboratories. Therefore, we implemented and described an exemplary web application that supports an efficient process-oriented management of NGS-related meta data.

One important advantage of our REST-based prototype is the increasing number of RESTful applications and clients that can be easily interfaced in future, for example the biobank software *OpenSpecimen* [9]. HL7 FHIR is an important standard for RESTful medical software applications that makes RESTful service-oriented architectures increasingly attractive for the data management of a variety of “non-standard” clinical data.

Due to project time constraints, there are some limitations. Today, the prototype is not yet in routine use. We collected positive reactions from future users and measured decreased processing time, but no systematic study of the effects of the new solution has been conducted. Future work will also concentrate upon more sophisticated statistical reports for the data.

References

- [1] Seth-Smith HMB et al. Evaluation of Rapid Library Preparation Protocols for Whole Genome Sequencing Based Outbreak Investigation. *Front Public Health*. 2019;27(7):241.
- [2] Kim KU et al. Comparison of functional gene annotation of *Toxascaris leonina* and *Toxocara canis* using CLC genomics workbench. *Korean J Parasitol*. 2013;51(5):525-30
- [3] Protsyuk IV et al. Shared bioinformatics databases within the Unipro UGENE platform. *J Integr Bioinform*. 2015;12(1):257.
- [4] Hong EL et al. Principles of metadata organization at the ENCODE data coordination center. *Database* 2016. 2016;baw001.
- [5] Figma Design. Figma: the collaborative interface design tool, Retrieved October 14, 2019 from <https://www.figma.com/>.
- [6] IEEE Computer Society Software Engineering Standards Committee. IEEE Recommended Practice for Software Requirements Specifications, IEEE Standard 830-1998, 1998.
- [7] Cervone HF. Understanding agile project management methods using scrum. *OCLC Syst Serv*. 2011;27(1):18-22.
- [8] Hasibuan A et al. Design and implementation of modular home automation based on wireless network, REST API, and WebSocket, In: *Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) - IEEE 2015*. 2015; 362–367.
- [9] McIntosh LD et al. caTissue Suite to OpenSpecimen: Developing an extensible, open source, web-based biobanking management system. *J Biomed Inform*. 2015; 57:456-464.

Providing an Integrated Access to EHR Using *Electronic Health Records* Aggregators

Belen PRADOS-SUAREZ^{a,1}, Carlos MOLINA^b and Carmen PEÑA-YAÑEZ^c

^aUniversity of Granada, Granada, Spain

^bUniversity of Jaen, Jaen, Spain

^cSan Cecilio Hospital, Granada, Spain

Abstract. The integration of health data systems is an open problem. Most of the active initiatives are based on the use of standards, each one proposed for a concrete use, without integrating other needs or standards allowing on homogenous use. We propose an alternative to get an unified view of health related data, valid for several uses, that can integrate heterogeneous sources. The proposal set the framework to integrate the developments made so far to automatically learn the extraction and conversion of the data. All the sources are integrated under a single access point. We present the main concepts of *EHR_{agg}* as a middleware between systems that can incorporate several sources giving an unified access following the FAIR principles.

Keywords. Health Data, Interoperability, Data Mining, Natural Language Processing

1. Introduction

Nowadays the need for interoperability between Hospital Information Systems is obvious, as it enables patient mobility, not only geographically, but also between medical services and health care providers.

Many efforts have been made to achieve this *interoperability*, starting with standardization initiatives (e.g. [9,10]). The main issues with these standards are two: first, there is no agreement about which one to choose; and second, it is going to take long before all the systems comply with these standards and become effectively interoperable, due to the huge efforts required to adapt current ad hoc hospital information systems.

This is a problem that also affects data *accessibility*, especially in health research, where data science offers very promising and useful tools ([1,2,3]), but has limited data to work with. This is due to the lack of integrated and standardized datasets for health data. Many initiatives to improve the quality and number of sources for research have been proposed (e.g [4,5,6]). For example, both USA ([7]) and Europe ([5]), have developed catalogues where researchers can add references to datasets and look for others, but

¹Corresponding Author: Belen Prados-Suarez. University of Granada. e-mail: belenps@ugr.es. Partially supported by PGC2018-096156-B-I00 *Recuperación y Descripción de Imágenes mediante Lenguaje Natural usando técnicas de Aprendizaje Profundo y Computación Flexible* of the Ministerio de Ciencia, Innovación.

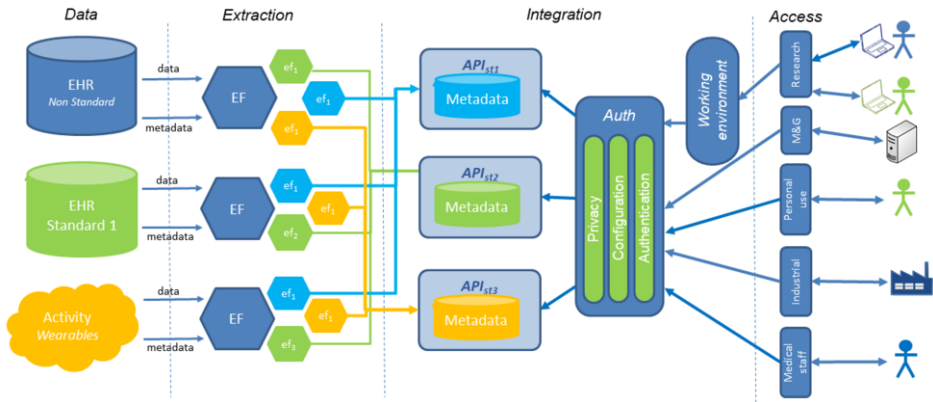


Figure 1. Architecture for EHR_{agg} systems

they need to adapt the data or methods to work with each of the datasets. As Schulz et al. indicate ([8]), although several standards have been developed, they are not generic enough to be valid for different uses.

Moreover, there is a third variable to add to the current landscape. New personal devices are emerging every day, which gather very useful information about people's way of life, but they are developed by private companies under different ad-hoc implementations, and completely disconnected from the rest of the systems and datasets.

In sum, we are faced with the following circumstances: (1) several sources of information, not only from medical institutions, that should be integrated; (2) different access needs with distinct requirements (personal use, medical use, research purposes,...), especially regarding privacy protection; and (3) the need for a homogeneous access point for all this data, with adaptation capabilities (to fit new systems, standards and needs), but without the need to transfer the ownership of the data.

The main idea is to build what we call an *EHR Aggregator* (EHR_{agg}) that integrates the developments made so far (and upcoming ones) to automatically learn how to convert current information systems into standard systems. This system also includes a layer that operates as a homogeneous access point to the information stored in all the sources, without necessarily having to transfer the data, but enabling its processing.

Here we present the concept of EHR Aggregator and the conditions that each element should comply with in section 2. Our purpose here is to present the general structure. In section 3 we put forward our conclusions about how it aims to solve these issues.

2. General Structure

In this section we cover the concept of Electronic Health Record Aggregator. We first present the idea and the general structure, while the following subsections are dedicated to describing each of the elements in the structure.

An *EHR aggregator* (EHR_{agg}), in Figure 1, is a system that acts as middleware between EHR systems and others with health-related data (e.g. activity, sleep analytics, etc.) and the users that want to access the information (which can be people or other systems). It provides a unified view of the underlying Health Data Systems, in the *Data*

Layer, by learning automatically, and in the *Extraction Layer*, by converting the information stored within. This uniform access point allows the user to query and retrieve information from all of the systems without adapting the access to the architecture of each one, which is possible through the *Integration Layer*. This layer acts as a unique access point to all the collected data for external applications. To improve reusability, there are platforms in this *Access Layer* that collect and update the generic functions to facilitate the implementation of external applications. Each of these layers are discussed in the following sections.

2.1. Data Layer (DL)

One of the main elements of the structure are the sources of information. From each one we have to obtain not only the data, but also the metadata. We then need to convert or translate this information in order to integrate it into the aggregator.

Depending on the type of data, we can use different functions to perform this conversion. For example, the functions to convert metadata may be different from those to convert data, and those used for images may be different from those used for medical analysis information. Each of these functions are called *extraction function* (*ef*).

During this process it is essential to consider the quality of the conversion. In a perfect situation, we will translate the information without any loss of data or metadata. In this case we will consider a value $c_i^j = 1$, indicating a perfect conversion. However, in most of cases we will not have perfect conversions, so we shall indicate this using a value in the $[0, 1]$ range. The nearer to 1 the value, the better the conversion is, and the less information is lost.

2.2. Extraction Layer (EL)

The most critical part is to retrieve the data from the source systems and transform it so that the API (application programming interface, see next section) can access all the information. The components in charge of this, called *ef* (extraction functions), transform the information stored in the source systems, to allow the API to work with it.

If the source is based on a health standard (e.g. [9,10]), the conversion may be direct but not always is so ([8]). Most of the current health data systems are not completely based on standards, so we need to adapt the data. So far, great advances have been achieved in standards compliance. However, system standardization, in most situations, is very complicated or, under the best circumstances, is costly in time and effort. This is why we propose here to get the most of these advances by using methods that automatically learn how to perform the conversion from any system to one of the standards and between the standards themselves.

This automatic learning process allows us to (1) directly integrate systems that already comply with one standard; (2) reuse the learned transformations to apply them in the integration of other non-standard compliance systems, and (3) keep open to future standards, sources or advances in integration or interoperability.

Depending on the type of data, we may define different *ef* functions. For non-standard systems, the *ef* need to learn the structure of the data (and metadata) and transform it to be used by the *EHR_{agg}*. Most EHR systems are based on semi-structured documents and we can find several proposals in the literature using Text Mining to extract

data from EHR (see [11]). Other approaches are based on Natural Language Processing (NLP) ([12,13,14]). These processes can be applied to extract not only data but also metadata. In both cases, the proposed solutions can be tested to obtain information regarding confidence in the conversion, and the best algorithm for each concrete type of data will be chosen. This value, normally in $[0, 1]$, provides the user with information regarding the confidence of the data offered by the aggregator. We can define the *ef* function combining several proposals according to the data to be transformed.

2.3. Integration Layer (IL)

The *IL* has three components. First, the *API* itself, with the methods to access and/or retrieve the data from *EL* and providing access to the functions and the working environment. The *authentication* module, which controls and registers the accesses, configures privacy restrictions, and ensures compliance with regulations. The last one is the *working environment*, where the access platforms can process all the data without having to leave the system, and avoiding the transfer of the data to other layers.

Through the *API*, the user has access to the information stored in each and all of the data sources integrated in the *EHR_{agg}*. It maximizes accessibility to information, with the only restrictions being those imposed by regulations and privacy protection rules. Using this *API* to develop their applications, users can run queries over all the data and metadata available in a homogeneous way, regardless of where it comes from or how it is stored. It is possible because the *API* makes use of the extraction functions to retrieve the data.

The second element is the *Authentication* module. It has the following functions: To register all of the data retrievals, regardless of the *API* used (*authentication module* in *Auth* block (*AB*)); to control privacy protection restrictions imposed by the source that provides the data or according to how the data will be used (*privacy* in *AB*); to adapt the access point to different regulatory frameworks (*regulation configuration* in *AB*).

Finally, we have the *working module*, that enables direct processing of data without really having the data. The idea is to offer a work space where, for example, researchers can use the *API* to launch their methods and perform calculations on the data, without retrieving the data or transferring it from the source institution.

2.4. Access Functions (AF)

The access layer is used to run queries in the system. Since we are referring to a large-scale system, we have taken into account that there will be different access requirements, depending on how it will be used. For example, research activities may require anonymized data and statistical methods, while a medical EHR system needs access to detailed information of the EHR of a concrete patient homogeneously, when the data in the original sources is spread out and fragmented. Therefore, as a starting point, we have identified at least five types of data use with distinct needs. These are: personal use of data, medical use, use for research, for management and governance purposes, and industrial use. This is not a closed classification, since new needs may arise in the future.

Here we propose to build an *access platform* per use or purpose where, in a similar way to software as a service (SaaS) platforms, the most frequent operations will be implemented for each given use. This way, as shown in figure 1, they can be reused by *external functions*, applications and systems of users with similar needs.

Finally, *external functions* can only access the system through the *access platforms*, which in turn have access from the *working module* or directly through the *API*. Let us remark that in any case, all of the access points pass through the *authentication module*.

3. Conclusions

With the EHR_{agg} we propose to address the interoperability and accessibility problem using the same pragmatic approach: instead of trying to have all the systems agree with the same standard, we propose a translation between standards, and of systems to any standard, reducing effort and time. It includes an integration layer that acts as a single access point, offering an unified view of the underlying data sources. This layer also ensures access control, privacy protection and adaptation and compliance with regulations. In addition, it has an access layer especially designed to facilitate the development of external access functions, based on the principle of reusability.

It is an ambitious proposal that aims to centralize current efforts to accelerate understanding between health-related systems. Here we present the general structure of the *EHR aggregator*, establishing the first steps to build a unified framework, but there is still a lot of work to be done. Some of the current and future lines of work focus on the study of access time requirements, feedback on the *ef* used to improve the system, the classification of access functions and their assessment to be integrated in the access platforms. Nevertheless, we believe this practical and open approach can give great results in the short and medium term.

References

- [1] Xu Lina, Simjanoska Monika, Koteska Bojana, et al. What Clinics are Expecting From Data Scientists? A Review on Data Oriented Studies Through Qualitative and Quantitative Approaches *IEEE Access*. 2019;7:641-654.
- [2] Health EU . Health EU – Human avatars to prevent and cure diseases <https://www.health-eu.eu/> 2019.
- [3] European Union . Future Health <https://www.futurehealtheuropa.eu/> 2016.
- [4] National Institutes of Health (USA) . NIH Strategic plan for data science (<https://datascience.nih.gov>)
- [5] Oliveira José Luís, Trifan Alina, Silva Luís A. Bastião. EMIF Catalogue: A collaborative platform for sharing and reusing biomedical data *Int. J. Med. Informatics*. 2019;126:35 - 45.
- [6] NCBC (USA) . Informatics for Integrating Biology and the Bedside (<https://www.i2b2.org>)
- [7] Margolis Ronald, Derr Leslie, Dunn Michelle, et al. The National Institutes of Health's Big Data to Knowledge (BD2K) initiative *J. Am. Med. Inform. Assoc.*. 2014;21:957–958.
- [8] Schulz Stefan, Stegwee Robert, Chronaki Catherine. *Standards in Healthcare Data*:19–36. Springer 2019.
- [9] standard HL7. <http://www.HL7.org> 2011.
- [10] ISO-13606 . ISO 13606: Electronic health record communication 2008.
- [11] Simmons Michael, Singhal Ayush, Lu Zhiyong. *Text Mining for Precision Medicine: Bringing Structure to EHRs and Biomedical Literature to Understand Genes and Health*:139–166. Springer 2016.
- [12] Dreisbach Caitlin, Koleck Theresa A., Bourne Philip E., Bakken Suzanne. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data *Int. J. Med. Informatics*. 2019;125:37 - 46.
- [13] Kaur Rajvir. *A comparative analysis of selected set of natural language processing (NLP) and machine learning (ML) algorithms for clinical coding using clinical classification standards*. PhD thesis 2018.
- [14] Chen Liang, Song Liting, Shao Yue, Li Dewei, Ding Keyue. Using natural language processing to extract clinically useful information from Chinese electronic medical records *Int. J. Med. Informatics*. 2019;124:6 - 12.

Retrieving and Analyzing Hospital Service Suspensions from Regional Healthcare Insurance Claims Data

Jumpei SATO^{a,1}, Kazutoshi UMEMOTO^a, Kazuo GODA^a,
Masaru KITSUREGAWA^a and Naohiro MITSUTAKE^b

^a*Institute of Industrial Science, The University of Tokyo*

^b*Institute for Health Economics and Policy*

Abstract. The geographical imbalance of the healthcare workforce is a social problem in Japan. Except for big cities, hospitals have difficulties in securing a sufficient workforce to offer healthcare services stably. For local government, hospital service suspensions are potentially an essential indicator to figure out the capacity of the regional healthcare supply. This paper proposes an algorithm that automatically identifies and classifies hospital service suspensions from insurance claims data, based on periodicity and similarity. To verify the effectiveness, we have applied the algorithm to the insurance claim dataset, which has been provided 91 regional public insurers in Japan. The case studies have confirmed that the proposed algorithm has presented an evidential picture of hospital service suspensions, which is potentially useful to understand the actual capacity of healthcare service supply in regions.

Keywords. Hospital service suspension, insurance claims data

1. Introduction

The geographical imbalance of the healthcare workforce is a common problem in many countries. Japan is no exception [1-2]. To tackle this problem, the Ministry of Health, Labor and Welfare of Japan has been executing several healthcare policies from 2008 [3]. As a result, the number of health workforce (especially physicians) has been increasing. However, a recent study showed that the supply capacity of the health workforce was still imbalanced between urban and rural areas in Japan [4]. In addition, this study showed that the supply capacity of the health workforce was decreasing in rural areas. Except for limited big cities, hospitals have difficulties in securing a sufficient workforce to offer healthcare services stably. Hence, understanding the balance supply capacity and healthcare demands is crucial for local governments to design their healthcare policies effectively.

To figure out the capacity of the regional healthcare supply, hospital service suspensions are potentially an essential indicator for local governments. In this paper, we present our idea of identifying hospital service suspensions from healthcare insurance claims data. Japan has established the universal health insurance system since 1961 [5];

¹ Corresponding Author, Jumpei Sato, Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-Ku, Tokyo, Japan; E-mail: jsato@tkl.iis.u-tokyo.ac.jp

all the medical services are covered by the public healthcare insurance programs. The healthcare insurance claims have the great potential to reveal the hospital service suspensions thoroughly and precisely.

In this study, we propose an algorithm that automatically detects and classifies the hospital service suspensions from a large amount of insurance claims data, based on periodicity and specificity and present case studies to demonstrate the effectiveness of the algorithm. The remainder of the paper is organized as follows. Section 2 describes the proposed algorithm, Section 3 presents case studies using four and half years of insurance claims dataset, and finally, Section 4 presents conclusions and discussions.

2. Method

The algorithm proposed in this paper detects and classifies hospital service suspensions from healthcare insurance claims data. The algorithm can be roughly divided into the following two steps: Step1; scans the given healthcare insurance claims data and then identifies services suspensions of each hospital based on the daily occurrence of medical services, and Step2; classifies service suspensions based on the periodicity and the similarity of suspension occurrences and the heuristic rules. The following two sections describe these steps in detail.

2.1. Identify hospital service suspension from the healthcare insurance claims data

Healthcare insurance claims data (standardized in Japan) does not directly contain the information of hospital service suspension. Instead, the claims data contain detailed information of medical service provision from hospitals to patients. That is, each claim records a date, a hospital identifier, a patient identifier and one or more medical service codes. Firstly, the algorithm aggregates a large amount of healthcare insurance claims to predict off (non-consultation) days of each hospital. Specifically speaking, the algorithm calculates, from the insurance claims, a service provision function for each hospital, $y(t)$, which denotes the number of patients that the hospital consulted on a date t . Obviously, the date tt satisfying $y(t) = 0$ is highly likely to be an off day for the hospital. Our algorithm thoroughly scans the healthcare insurance claims to list all the off days for each hospital; if four and more consecutive days are off for a concerned hospital, these off days are recognized as a service suspension².

2.2. Classify hospital service suspensions

The proposed algorithm classifies the hospital service suspensions into the following four types based on the periodicity and the similarity of suspension occurrences and the heuristic rules: (a) apparent service suspensions before hospital open, (b) apparent service suspensions after hospital close, (c) anormal service suspensions, and (d) regular service suspensions (planned, including weekends and holidays).

First, by the heuristic rules, the algorithm classifies the suspensions into the before hospital open or the after hospital close. If the duration of the suspension is over a month

² In Japan, many organizations including hospitals are off on Saturday and Sunday, and national holidays are often placed on Monday. After intensive exploration, we came to define four and more consecutive off days as a suspension to distinguish such weekend off days.

and started from the first data of given claims data, the algorithm classifies the suspension as the before hospital open. Likewise, if the duration of the suspension is over a month and ended on the last date of the given claims data, the algorithm classifies the suspension as the after hospital close.

Next, based on the periodicity and the similarity of suspension occurrences, the algorithm classifies the remaining suspensions into the anormal or regular service suspensions. Generally, scheduled hospital suspensions would occur at regular intervals (weekly, annually)². Therefore, the algorithm calculates the periodicity $pw(t)$ by using the trends and cycles model [7] as follows:

$$y(t) = g(t) + s(t) + \varepsilon_t, \quad pw(t) = \begin{cases} y_{fitted}(t), & (t \mid y(t) = 0) \\ 0, & (t \mid y(t) \neq 0) \end{cases} \quad (1)$$

Where $y(t)$ is the time-series data calculated in Step1, $g(t)$ denotes a constant up/down trend, and $s(t)$ denotes the weekly/annually periodic changes. ε_t denotes error terms. $y_{fitted}(t)$ represents a fitting result of the trends and cycles model.

Furthermore, in Japan, many hospitals schedule their day off around August 14 (called the “Obon” holidays), despite not national holidays. Such as the “Obon” holidays, the suspensions when many hospitals were suspended similarly can be considered as scheduled suspension. Hence, the algorithm calculates the similarity $sw(t)$ referring to the idf (inversed document frequency) index, which can measure whether a suspension is similar (common) or rare:

$$sw(t) = \begin{cases} \log_2 \left(\frac{\# \text{ of suspended hospitals on each date: } t}{\# \text{ of total hospitals}} \right), & (t \mid y(t) = 0) \\ 0, & (t \mid y(t) \neq 0) \end{cases} \quad (2)$$

Our algorithm classifies the remaining suspension, which contains the four and more consecutive days that satisfy $pw(t) + sw(t) > 16$, as the anormal service suspension. Finally, the algorithm classifies the remaining suspension as the regular service suspension.

3. Result

In our study, we used the insurance claims dataset of 91 regional public insurance providers in the Mie prefectural area in Japan. This dataset contained 71.53 million claims (1.676 billion records) from March 2013 to November 2017.

3.1. Case study1: detecting and classifying service suspensions

First, we applied the proposed algorithm to the dataset. Figure 1 shows an example of the visualization result of the suspended dates for each hospital. In Figure 1, the horizontal axis represents date t , the vertical axis represents each hospital, and each column shows the number of patients who provided medical services on each date. The no service provided dates are shown in colored. As shown in Figure 1, in many hospitals, the no service provided dates appeared at a cycle of a week. In addition, early May, which

has many national holidays, “Obon” holidays and New Year’s holidays showed that many hospitals were suspended at the same time.

Table 1 shows the details of the number of suspends detected and classified by the algorithm. The algorithm identified 12,813 service suspensions in total, which were classified into (a) 12,217 regular service suspensions, (b) 523 anormal service suspensions, (c) 26 hospital opens, and (d) 47 hospital closes. The number of service suspensions the highest in the urban areas (populations are $\geq 100K$), second in the medium areas (populations are $<100K$ and $\geq 10K$), third in the rural areas (populations are $<10K$) for each suspension type.

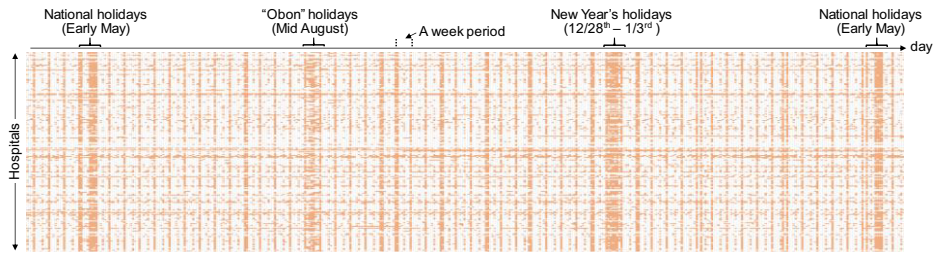


Figure 1. Example of the visualization of the suspended days (the no service provided dates are colored).

Table 1. Details of the number of suspend detected and classified by the algorithm.

	Regular service suspension	Anormal service suspension	Hospital open	Hospital close
Urban areas	8,741	428	20	32
Medium areas	3,358	89	6	13
Rural areas	118	6	0	2
Total	12,217	523	26	47

Figure 2 shows the average duration of the regular and anormal service suspensions. The average duration of the regular service suspensions was 5.37 days in the urban areas, being significantly longer than that of the medium areas (5.21 days ($p<0.01$)) and the rural areas. (4.81 days ($p<0.05$)). On the other hand, there was no significant difference in anormal service suspensions (the urban areas: 8.51 days, the medium areas: 10.04 days, the rural areas: 8.17 days). Figure 3 shows the average frequency of the regular and anormal service suspensions. The average frequency of the regular service suspensions was 5.90 times in the rural areas, being significantly lower than that of the urban areas (11.52 times ($p<0.01$)) and the medium areas (11.82 times ($p<0.05$)). On the other hand, there was no significant difference in anormal service suspensions (the urban areas: 2.30 times, the medium areas: 1.93 times, the rural areas: 2.00 times).

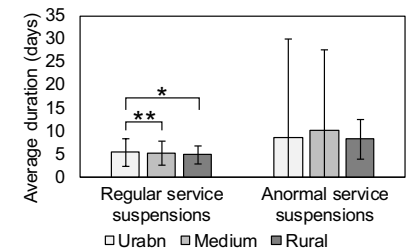


Figure 2. The average duration of the service suspensions (*: $p<0.05$, **: $p<0.01$).

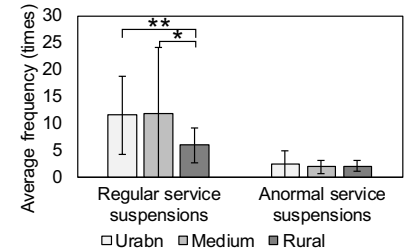


Figure 3. The average frequency of the service suspensions (*: $p<0.05$, **: $p<0.01$).

These results show that regular service suspensions tend to shorter and fewer in rural areas. This trend indicates that the number of hospitals in rural areas is limited, and there are no alternative hospitals in their regions.

3.2. Case study2: the relationship between the service suspensions and the demand behavior

Next, we analyzed the relationship between the service suspensions and the demand behavior of patients. In the dataset, 55 hospitals provided consecutive dialysis treatment to 5,450 patients. In these hospitals, regular and anormal service suspension did not occur. From this result, we confirmed that a stable supply system was established in the hospitals that provided continuous dialysis. On the other hand, 144 hospitals provided consecutive rehabilitation treatment to 29,513 patients. Of these, anormal service suspension occurred in 12 hospitals. 412 patients affected by the service suspensions, of which 406 patients were not receiving rehabilitation during the suspensions and 6 patients were receiving rehabilitation from other hospitals.

This result shows that the anormal service suspensions could occur even in the hospitals that provide consecutive treatments, excepting the hospitals that provide treatments, which have a high impact on life support such as the dialysis.

4. Discussion and Conclusion

In this study, we have proposed an algorithm for detecting and classifying hospital service suspensions from healthcare insurance claims data, and then presented the case studies, which confirm that the algorithm has effectively presented an evidential picture of hospital service suspensions. The present study has been performed with the dataset provided by 91 regional public insurers in Japan. We would like to extend the study to other regions including metropolitan areas.

The next possible challenge is to reveal the resilience of the regional healthcare system, such as how neighboring hospitals compensate suspended healthcare services.

References

- [1] G Dussault, MC Franceschini, Not enough there, too many here: understanding geographical imbalances in the distribution of the health workforce, *Hum Resour Health* 4(1):12 (2006). doi:10.1186/1478-4491-4-12.
- [2] P Zurn, MR Dal Poz, B Stilwell, O Adams, Imbalance in the health workforce, *Hum Resour Health* 2(1):13 (2004). doi:10.1186/1478-4491-2-13.
- [3] Ministry of Health, Labour and Welfare, The physician uneven distribution measures. Available From: <https://www.mhlw.go.jp/file/06-Seisakujouhou-10800000-Iseikyoku/0000194394.pdf> (Japanese).
- [4] K Hara, T Otsubo, S Kunisawa, Y Imanaka, Examining sufficiency and equity in the geographic distribution of physicians in Japan: a longitudinal study, *BMJ Open* 7(3) (2017), e013922. doi:10.1136/bmjopen-2016-013922.
- [5] Ministry of Health, Labour and Welfare, Annual Health, Labour and Welfare Report 2011. Available from <https://www.mhlw.go.jp/wp/hakusyo/kousei/11/dl/01-02.pdf> (Japanese).
- [6] Ministry of Health, Labour and Welfare, Various information of medical fee. Available from <http://www.iryohoken.go.jp/shinryohoshu/> (Japanese).
- [7] AE Harvey, S Peters, Estimation Procedures for Structural Time Series Models, *Journal of Forecasting* 9 (1990), 89-108. doi:10.1002/for.3980090203.

Secure Collapsing Method Based on Fully Homomorphic Encryption

David NIYITEGEKA ^a, Reda BELLAFQIRA ^a, Emmanuelle GENIN ^b and
Gouenou COATRIEUX ^{a,1}

^aINSERM UMR 1101 LaTIM, IMT Atlantique, Brest, France

^bINSERM, Univ Brest, EFS, CHU Brest, UMR 1078 GGB, F-29200 Brest, France

Abstract. In this paper, we propose a new approach for performing privacy-preserving genome-wide association study (GWAS) in cloud environments. This method allows a Genomic Research Unit (GRU) who possesses genetic variants of diseased individuals (cases) to compare his/her data against genetic variants of healthy individuals (controls) from a Genomic Research Center (GRC). The originality of this work stands on a secure version of the collapsing method based on the logistic regression model considering that all data of GRU are stored into the cloud. To do so, we take advantage of fully homomorphic encryption and of secure multiparty computation. Experiment results carried out on real genetic data using the BGV cryptosystem indicate that the proposed scheme provides the same results as the ones achieved on clear data.

Keywords. GWAS, security, homomorphic encryption, cloud computing

1. Introduction

Nowadays, a large amount of genetic data is available enabling the better understanding of the human genome, especially by means of genome-wide association studies (GWAS). The main objective of GWAS is to identify genetic variants associated with some diseases. These studies are mainly based on the statistical analysis of genomic data shared between a genomic research unit (GRU) who possesses genetic variants of affected individuals (cases) and a genomic research center (GRC) who has genetic variants of unaffected individuals (controls). Today, GWAS aims at taking advantage of cloud computing which allows data owners to store and process large amount of data. However, genetic data externalization raises up many questions in terms of data privacy and confidentiality. Indeed, with outsourcing an owner loses the control over his data. Even if a service level agreement has been signed with the cloud service provider (CSP), the user has actually no other choice than trusting him. In addition, the human genome is a very sensitive. It is unique to its owner and can give information about his or her relatives, from a clinical and behavioral point of view. Thus, there is a need for protecting genetic data during their storage or processing in the cloud. In this work, we are interested in securing the

¹Corresponding Author: Gouenou Coatrieux, IMT Atlantique, Inserm, UMR 1101 Latim, 29238 Brest Cedex, France; E-mail: gouenou.coatrieux@imt-atlantique.fr.

collapsing method [1], a case-control study GWAS, with as objective to test whether the proportions of individuals with rare variants in case sample and control sample differ.

Several solutions have been proposed so as to perform privacy-preserving GWAS. The first ones were based on data anonymization techniques such as k -anonymity, but have been shown vulnerable to many attacks [2]. Methods based on differential privacy (DP) [3] overcome these issues. DP adds a random noise to real data so as to ensure individuals' privacy but at the same time reduces data usability. Homomorphic encryption (HE) and secure multiparty computation (SMC) are also used. HE makes possible performing linear operations on encrypted data without decrypting them with the guarantee that the decrypted result equals to the one computed on clear data [4]. SMC allows several parties to jointly compute a function over their data while keeping them private [5]. There exist several approaches based on secret sharing, oblivious transfer and HE [6]. Many HE and SMC based methods have been proposed to secure GWAS statistical algorithms used such as χ^2 -statistic without compromising data usability [5,7,8]. In this paper, we present the first scheme that allows the secure computation of collapsing method based on the logistic regression model. It takes advantage of fully homomorphic encryption, and especially of the BGV cryptosystem which allows encrypting and process blocks of data, in combination with SMC. Contrarily to the previous schemes, it considers that all user data are outsourced and only returns to the users whether the test is significant or not making our solution more secure.

The rest of this paper is as follows. Section 2 comes back on some cryptographic preliminaries and on the collapsing method. Section 3 presents our secure collapsing method. Experimental results and conclusion are given in Section 4 and 5, respectively.

2. Cryptographic and Collapsing method preliminaries

2.1. Homomorphic encryption and secure multiparty computation

Homomorphic encryption (HE) allows performing linear operations (e.g., "+", "×") over encrypted data. Formerly, if E and D denote the encryption and decryption functions, and K_p and K_s are the public and the secret keys respectively, then the homomorphic property states: $D(E(m_1, K_p) \star E(m_2, K_p), K_s) = m_1 \circ m_2$ where: \star and \circ are two operators in the encrypted and the clear/decrypted domains; m_1 and m_2 are two plain-texts. There are three types of HE cryptosystems: partially HE cryptosystems (allow computing additions or multiplications unlimitedly); somewhat HE cryptosystems (allow a limited number of additions and multiplications); and, fully HE cryptosystems (allow unlimited number of additions and multiplications). In this work, we opted for BGV [9], a fully HE cryptosystem because of its homomorphic and batching properties. Batching consists in encrypting multiple messages in one with the capacity to process each message. These properties are of importance in our scheme.

2.2. Collapsing method

This method involves collapsing genotypes across variants in a specific gene and applying a statistical test such as logistic regression. It is a powerful tool for analyzing rare variants. Let us consider a sample of N individuals constituted of N_1 cases and N_2 con-

trols. Let Y_j be the disease status: Y_j equal 1 if an individual j is a "case" or 0 if he/she is a "control". To perform the test, k sites of the studied gene where the variants of interest exist are chosen by GRC and GRU. After this selection, a variable X_i is defined for each individual. It equals to 1 if the individual i has at least one variant allele on any of the k sites or 0 otherwise. The detection of the association consists in testing if the proportion of individuals in cases and controls differ. To do so, a logistic regression model [10] can be used where the outcomes are $\{Y_j\}_{j=1\dots N}$ and the predictors are the variables $\{X_j\}_{j=1\dots N}$. The simplified logistic regression model is defined as $\text{logit}(p) = \ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X$, where p is the probability that Y_j occurs and $\{\beta_l\}_{l=0,1}$ are the regression coefficients. These ones are estimated using the log of the maximum likelihood function such that

$$l(\beta) = \log(L(\beta)) = \sum_{i=1}^{N_1+N_2} y_i(\beta_0 + \beta_1 x_i) - \sum_{i=1}^{N_1+N_2} \ln(1 + \exp(\beta_0 + \beta_1 x_i)) \quad (1)$$

To conduct the collapsing method, two hypothesis tests are considered. The null hypothesis (H_0) states that the studied gene is not associated to the disease. In that case, β_1 equals to 0 and $\text{logit}(p) = \beta_0$. The alternative hypothesis (H_1) informs that there is an association. β_1 is not null and the logit model remains $\text{logit}(p) = \beta_0 + \beta_1 X$. Let us define a , b , c and d as the number of individuals that: have at least one variant allele on any of the k sites in controls; have at least one variant allele on any of the k sites in cases; have non variant allele on the k sites in controls and have non variant allele on the k sites in cases, respectively. In that case, (1) becomes

$$l(\beta) = d\beta_1 + b(\beta_1 + \beta_2) - [(a+b)\ln(1 + \exp(\beta_1 + \beta_2))] - [(c+d)\ln(1 + \exp(\beta_1))] \quad (2)$$

This function has a global maximum only if a , b , c and d are different from 0. The maximization of (2) under H_1 gives us $\hat{\beta}_{H_1} = (\hat{\beta}_0, \hat{\beta}_1) = (\ln(\frac{d}{c}), \ln(\frac{b*c}{a*d}))$ while under H_0 , it leads to $\hat{\beta}_{H_0} = \hat{\beta}_0 = \ln(\frac{b+d}{a+c})$. After estimating regression coefficients, the statistic test result is given by $Stat = 2(l(\hat{\beta}_{H_1}) - l(\hat{\beta}_{H_0}))$. This value being distributed according to the χ^2 distribution with 1 degree of freedom ($\chi^2(1)$), if $Stat > \chi^2(1)$, H_0 is rejected, otherwise, we can not decide whether the gene is associated to disease or not.

3. Secure collapsing method

Our framework considers three entities: a Genetic Research Unit (GRU) who owns N_1 cases; a Genomic Research Center (GRC) who has N_2 controls and a Cloud Service Provider (CSP). We assume that: *i*) all data are BGV encrypted and stored by CSP, GRU and GRC possess their own pair of keys, respectively i.e., (K_p^U, K_s^U) and (K_p^C, K_s^C) and *ii*) GRU and GRC ask CSP to perform collapsing method on their data. CSP is considered as "honest but curious". It follows the processing steps but may try to infer information about GRU and GRC data. Herein, we only focus on protecting data confidentiality and individual privacy. Other data threats are out of the scope of this paper due to space limitation. Because the collapsing method is based on some nonlinear operations that cannot be achieved with BGV (e.g., division and $\ln(\cdot)$ function), a third party entity (TPE) is introduced. It has also a pair of key (K_p^T, K_s^T) . Case data of GRU and control data of GRC correspond to VCF files. To roughly sum up, one VCF file contains a table the lines

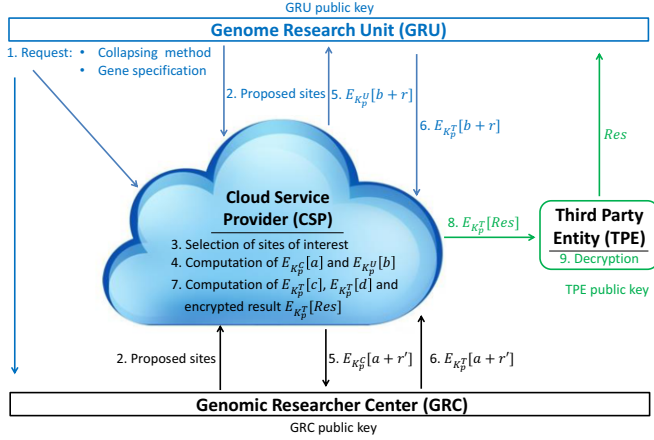


Figure 1. Different steps of our secure collapsing method

of which correspond to the variants and the columns to individuals. Each table element contains binary values. To benefit of the BGV batching property and of the fact it allows binary operations, one line is encrypted as a single message. For a test of M variants, we have two sets of encrypted vectors $\{E_{K_p^U}[v_n^{cas}]\}_{n=1\dots M}$ and $\{E_{K_p^C}[v_p^{con}]\}_{p=1\dots M}$. As shown in the Figure 1, our secure collapsing method is conducted in 9 steps as follows.

1. GRU sends a request to the CSP and GRC to initialize the test, specifying the gene he is interested in.
2. GRU and GRC choose their respective sites of interest and send them to CSP. They correspond to positions into the vectors $\{E_{K_p^U}[v_n^{cas}]\}_{n=1\dots M}$ and $\{E_{K_p^C}[v_p^{con}]\}_{p=1\dots M}$.
3. CSP selects the sites of interest in the stored samples of GRC and GRU and constructs two encrypted vectors of binary values $E_{K_p^U}[C_U]$ and $E_{K_p^C}[C_C]$ from $\{E_{K_p^U}[v_n^{cas}]\}_{n=1\dots M}$ and $\{E_{K_p^C}[v_p^{con}]\}_{p=1\dots M}$, respectively.
4. CSP computes $E_{K_p^C}[a] = E_{K_p^C}[\sum_k C_C(k)]$ and $E_{K_p^U}[b] = E_{K_p^U}[\sum_k C_U(k)]$.
5. CSP selects two random values r and r' , computes $E_{K_p^C}[r]$ and $E_{K_p^U}[r']$ and sends $E_{K_p^C}[a+r] = E_{K_p^C}[a] + E_{K_p^C}[r]$ to GRC and $E_{K_p^U}[b+r'] = E_{K_p^U}[b] + E_{K_p^U}[r']$ to GRU. This process corresponds to an additive data masking operation.
6. GRU decrypts $E_{K_p^C}[a+r]$ and GRC decrypts $E_{K_p^U}[b+r']$. Both of them re-encrypt these values using K_p^T and send the results to CSP.
7. CSP computes $E_{K_p^T}[N_1]$ and $E_{K_p^T}[N_2]$. With $E_{K_p^T}[a]$, $E_{K_p^T}[b]$, it calculates $E_{K_p^T}[c]$ and $E_{K_p^T}[d]$, and interacts with TPE to SMC computes $E_{K_p^T}[Stat]$ (see Section 2).
8. Thanks to BGV, CSP computes the encrypted sign $E_{K_p^T}[Res]$ of the encrypted difference $E_{K_p^T}[Stat - \chi^2(1)]$ and sends it to TPE.
9. TPE decrypts Res and sends the decrypted value to the GRU.

As stated in Section 2, the computation of $Stat$ requires the computation of $\ln(\cdot)$, a non-linear function. To secure it, our solution combines homomorphic encryption with an

original multiplicative data masking. It adds a noise that can be removed thanks to \ln property: $\ln(ab) = \ln(a) + \ln(b)$. Due to space limitation we cannot detail this approach.

4. Experimental results

Experiments were conducted on a genetic sample of 57 individuals (20 cases and 37 controls). These data are extracted from a real genetic database. Data are encrypted using the BGV cryptosystem implemented with HELib library. Our choice of HELib stands on the fact it optimizes the size of encrypted data to store in the cloud thanks to batching. We ran the proposed solution on a machine equipped with 4 GB RAM, Intel Core i5-5200U, 2.7GHz, running on Ubuntu 18.04 LTS. Execution time is about 15 minutes with no errors in the test results compared to the same tests conducted on clear data.

5. Conclusion

In this paper we have proposed the first privacy-preserving collapsing method using a logistic regression model. It takes advantage of fully homomorphic encryption, of SMC and multiplicative data masking in order to allow two entities to compute this test on encrypted data; data they store in the cloud without decrypting them. This solution is secure under the honest but curious adversarial model. Because our scheme makes no approximations, it achieves exactly the same results as working on clear data. Future works will focus on minimising computation complexity by exploiting parallel computing.

Acknowledgement: This work was granted to the Labex CominLabs and managed by the French ANR in the "Investing for the future" program, ref.ANR-10-LABX-07-01.

References

- [1] B. Li and S. M. Leal. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *The American Journal of Human Genetics*, 83(3):311–321, 2008.
- [2] Y. Erlich and A. Narayanan. Routes for breaching and protecting genetic privacy. *Nature Reviews Genetics*, 15(6):409, 2014.
- [3] S. Simmons, B. Berger, and C. S. Sahinalp. Protecting genomic data privacy with probabilistic modeling. In *PSB*, pages 403–414. World Scientific, 2019.
- [4] D. Niyitegeka, G. Coatrieux, R. Bellafqira, E. Genin, and J. Franco-Contreras. Dynamic watermarking-based integrity protection of homomorphically encrypted databases—application to outsourced genetic data. In *International Workshop on Digital Watermarking*, pages 151–166. Springer, 2018.
- [5] J. M. Bloom. Secure multi-party linear regression at plaintext speed. *arXiv preprint 1901.09531*, 2019.
- [6] D. W. Archer, D. Bogdanov, Y. Lindell, L. Kamm, K. Nielsen, J. I. Pagter, N. P. Smart, and R. N. Wright. From keys to databases—real-world applications of secure multi-party computation. *The Computer Journal*, 61(12):1749–1771, 2018.
- [7] A. Mittos, B. Malin, and E. De Cristofaro. Systematizing genome privacy research: a privacy-enhancing technologies perspective. *Proceedings on Privacy Enhancing Technologies*, 2019(1):87–107, 2019.
- [8] J. J. Sim, F. M. Chan, S. Chen, B. H. M. Tan, and K. M. M. Aung. Achieving GWAS with homomorphic encryption. *arXiv preprint arXiv:1902.04303*, 2019.
- [9] Z. Brakerski, C. Gentry, and V. Vaikuntanathan. (leveled) fully homomorphic encryption without bootstrapping. *ACM Transactions on Computation Theory*, 6(3):13:1–13:36, July 2014.
- [10] E. Y. Boateng and D. A. Abaye. A review of the logistic regression model with emphasis on medical research. *Journal of Data Analysis and Information Processing*, 7(4):190–207, 2019.

Semantic Querying of Hospital Data Using an Ontology-Based Model of Discharge Summaries and ICD 10

Ihssen BELHADJ^{a,1} and Thierry BOUDEMAGHE^a

^a *Department of Medical Information, Methods and Research (S.IM.M.E.R), CHU Nîmes, Univ Montpellier*

Abstract. This paper focusses on the question of knowledge-based modeling of discharge summaries for semantic requesting in the context of hospital reporting process. An ontology was designed and implemented to capture patient data and required expert knowledge. The modeling process was carried out manually using Protégé after an initial reverse engineering of the official format of discharge summaries. An OWL 2 model was built integrating discharge summaries ontology and ICD 10 hierarchy. The framework is operationalized using the OpenLink Virtuoso database system (RDF store), enabling SPARQL queries and basic reasoning features. The evaluation was performed by implementing a use case based on patient comorbidities.

Keywords. Ontologies, OWL, RDFS, RDF, ICD 10, Protégé, OpenLink Virtuoso

1. Introduction

Hospital stakeholders need business reports for medical activity analysis, decision-making and strategic planning of care. Although data analytics are valued when they are simple, concise and relevant, their production remains a complex task, requiring manipulation of referential data such as diseases and procedures that are coded using standard terminologies. ICD 10 codes [1], used in this context to express principal and secondary diagnosis, refer to detailed descriptions of disorders. These fine-grain codes are inappropriate for business reporting because they do not respond to the “summary” feature of hospital indicators. In order to overcome this issue, data scientists map ICD 10 codes to high-grain terms that are more relevant to describe care chains and in-hospital pathways. One example of such a situation is medical activity reporting according to the acuteness/chronicity of diseases or their anatomical location. Such reporting processes require concrete substantial intervention of domain experts to build code glossaries. The weaknesses of this process consist in: (1) the dependence on domain expert knowledge and points of view; (2) the deprecation of used codes that require verification over time; and (3) the lack of semantic mapping between different glossaries. This observation led us to consider that building sustainable reporting services cannot rely solely on data

¹ Corresponding Author, Ihssen BELHADJ, S.IM.M.E.R., Nîmes University Hospital, Pl Pr Robert Debré, 30000 Nîmes, France; E-mail: ihssen.belhadj@chu-nimes.fr.

management strategies. A better and simpler method should take a comprehensive approach covering the management and formalization of underlying knowledge. The objective of this work is to build and operationalize a knowledge-based model of discharge summaries that enables semantic queries regarding basic properties of disease terminology such as their hierarchical organization.

2. Material and methods

2.1. Discharge summaries data and ICD 10 hierarchy

Discharge summaries are created for each hospitalization [2]. Collected data cover administrative information such as date of entry and discharge, diagnosis information coded with ICD 10, procedures information coded using CCAM [3], and finally the *Groupe Homogène de Malade* (GHM), which is the equivalent of the diagnosis-related group. These data are uploaded monthly to a government agency named the *Agence Technique de l'Information sur l'Hospitalisation* (ATIH) to issue activity-based payments. Data are also stored in local databases for hospital reporting and evaluation studies. An official format of data is published by the ATIH providing a technical description of data to be transmitted: position, definition, type and sometimes range of possible values. Two types of data elements are distinguishable: referential data to be coded according to defined terminologies; and non-referential data that refer to elementary data type values such as integer, date or string.

This work focuses particularly on ICD 10, as it is the standard resource for disease coding. The strengths of this terminological resource are: (1) the amount of data coded with this resource; (2) its complete coverage; and (3) the hierarchical granularity of disease descriptions. The weaknesses of ICD 10 are: (1) some inconsistencies in chapters and subchapters as grouping nodes; (2) the lack of defining properties of terms; (3) insufficient semantic mapping procedures between different terms (i.e. etiology/manifestation) pertaining to its hierarchical nature; and (4) periodic updates possibly modifying the hierarchy and/or the meaning of terms [4].

2.2. Ontological modeling principles of discharge summaries using OWL 2

Ontologies provide a formalized and shared vision about domain concepts, properties and their semantic relations [1]. They provide a modeling framework to simultaneously express data and knowledge. Their usefulness consists in enabling reasoning features on complex assertions about properties and consistency checking. The ontological principles assumed in our modeling process are: (1) considering data elements in the official format as domain concepts expressed with “owl classes”, (2) discharge summaries data as instances expressed with ‘owl individual’, (3) elementary data types as ‘owl data properties’ and (4) referential data as ‘owl data object properties’. We started by reverse engineering data formats. This generated a raw OWL ontology containing the classes and object properties. This ontology was then refactored manually to add definition domains and ranges.

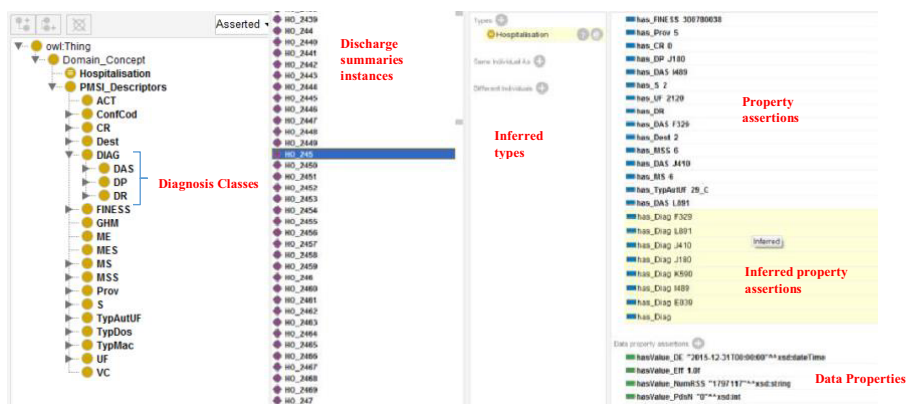


Figure 1 the OWL 2 model of discharge summaries description visualized using Protégé.

2.3. Intermediate model of the ICD 10

While terminology modeling has been addressed in detail [4,5], we focused on using the ICD 10 for its semantic information, e.g. its hierarchy, articulated with the data and allowing for semantic queries [6]. To this end, several issues had to be overcome: not only is the ICD 10 adapted for the French prospective payment system with extended codes, but this variant is also updated annually on March 1st, so an annual report should take into account two versions, not to mention multi-annual reporting.

Building on the current model of the ICD 10 released by the Organisation Mondiale de la Santé (OMS) [7] and expressed in SKOS [8], we designed an intermediate model by combining it with the French specific material:

- We retained only the hierarchical structure of the OMS (skos:broaderTransitive and skos:narrowerTransitive), stripping all other information (i.e. icd:classKind icd:CodingHint) apart from skos:preflabel, replaced by its French translation;
- Two elements were added as properties: the ICD 10 code (icdfr:code) formatted according to the French convention (code without dot); and the hierarchical level of the code (icdfr:nivH), starting from 1 for the chapter level. This information was grouped as properties of a resource <info>, <info> being itself a property of the ICD term resource;
- Starting from the 3 digit codes, we appended the French codes (e.g. including their extensions) and their hierarchical level. At this stage we had a complete French version of the ICD 10;
- Finally, we implemented a resource <version>, as a property of the <info> resource. <version> was designed with a start date, an end date and a version identifier. It was then possible to manage validity periods of the versions.

It is possible with Virtuoso to implement skos:broaderTransitive and skos:narrowerTransitive through “rule sets” associated to graphs [9]. We implemented the ICD 10 hierarchy using this functionality.

2.4. *Ontology population, data querying and reporting framework*

The reporting framework was implemented based on OpenLink Virtuoso, installed on a Windows 2012R2 Server running on an Intel Xeon E5 processor with 16 Go of RAM.

We uploaded around 340,000 triplets corresponding to the 70,000 discharge summaries of the Nîmes University Hospital for the first semester of 2019, which required querying both ICD 10 French version 2018 and 2019.

As an example, the following query retrieves in less than a second the number of discharge summaries by the ICD 10 codes of hierarchical level 2 their principal diagnosis is related to:

```
SPARQL DEFINE input:inference <http://localhost:8890/HIER>
PREFIX icdfr: <http://simmer.chu-nimes.fr/icd-fr-schema#>
PREFIX rss: <http://simmer.chu-nimes.fr/rss-schema>
SELECT DISTINCT ?ch ?lh (COUNT(DISTINCT ?NumRSS) AS ?NbRSS)
FROM <http://localhost:8890/RSS2019M06>
FROM <http://localhost:8890/REF>
WHERE
{
    ?x rss:hasValue_NumRSS ?NumRSS .
    ?x rss:has_DP ?d .
    ?d icdfr:info ?i .
    ?i icdfr:code ?c . ?i skos:prefLabel ?l .
    ?i icdfr:version ?v . ?v icdfr:ddv ?dd . ?v icdfr:dfv ?df .
    ?i skos:broaderTransitive ?ih OPTION (TRANSITIVE) .
    ?ih icdfr:code ?ch . ?ih skos:prefLabel ?lh .
    ?ih icdfr:nivH 2 .
    ?x rss:hasValue_DS ?ds
    FILTER
        (?dd <= ?ds && ?ds <= ?df)
}
GROUP BY ?ch ?lh
ORDER BY DESC(?NbRSS)
```

The combination of the `skos:broaderTransitive` property and the `icdfr:nivH` property ensures the determination of the level of aggregation. We used the R software to query the Virtuoso server and format and deliver the results.

3. Results and discussion

An ontology model of French discharge summaries was developed and expressed in OWL 2 formalism. This model is populated with large datasets available for requesting through SPARQL endpoint. SKOS model is used to implement the ICD 10 hierarchy. The whole framework supports hospital automatic reporting according to multiple criteria. Requesting data according to ICD 10 chapters or subchapters (intermediate codes that are not directly used in coding) was carried out taking advantages of reasoning features such as relation transitivity and “is-a”-property heritage.

This first step work shows the technical feasibility of the framework and lays out a path for sustainable hospital reporting services, through a sound methodology of data and knowledge management. Nevertheless, achieving such a purpose requires handling subsequent issues such as knowledge acquisition and knowledge sources consistency management.

To take into account the deeper properties of diseases will need the formalization of a knowledge-acquisition process. This raises the question of available disease knowledge sources and their evolution over time.

Protégé™ and OpenLink Virtuoso™ database systems are useful tools that help respectively to build the model and to operationalize the RDF database system. It would be useful if all Protégé reasoning features were implemented in Virtuoso™ and inversely if Virtuoso™ OWL model could be visualized via Protégé™ without ad-hoc developments.

4. Conclusion

Envisioning hospital reports through knowledge-engineering techniques has proven very fruitful, with already actionable results after this first round of exploration and development. This work enables us to identify the issues we need to explore further: enhance and refine the ontology of the discharge summaries in order to take into account not only the medical dimension but also other aspects (e.g. geographical origin of the patients and information about the hospitals); and expand the terminology modeling (e.g. procedure terminology) using the strategy we have defined to take into account the different versions.

Finally, we plan to study the national database using this approach in order to explore its capacity as an instrument for health resource usage analysis.

References

- [1] V. Bertaud-Gounot, R. Duvauferrier, and A. Burgun, Ontology and medical diagnosis, *Inform Health Soc Care*. **37** (2012) 51–61.
- [2] T. Boudemaghe, and I. Belhadj, Data Resource Profile: The French National Uniform Hospital Discharge Data Set Database (PMSI), *Int J Epidemiol*. **46** (2017) 392–392d. doi:10.1093/ije/dyw359.
- [3] B. Trombert-Paviot, A. Rector, R. Baud, P. Zanstra, C. Martin, E. van der Haring, L. Clavel, and J.M. Rodrigues, The Development of CCAM: The New French Coding System of Clinical Procedures, *Health Information Management*. **31** (2003) 2–11. doi:10.1177/183335830303100103.
- [4] J.J. Cimino, High-quality, Standard, Controlled Healthcare Terminologies Come of Age, *Methods Inf Med*. **50** (2011) 101–104.
- [5] A. Awaysheh, J. Wilcke, F. Elvinger, L. Rees, W. Fan, and K. Zimmerman, A review of medical terminology standards and structured reporting, *J Vet Diagn Invest*. **30** (2018) 17–25. doi:10.1177/1040638717738276.
- [6] S.W. Tu, S. Tu, L. Tennakoon, M. O'Connor, M. Connor, R. Shankar, and A. Das, Using an integrated ontology and information model for querying and reasoning about phenotypes: The case of autism, *AMIA Annu Symp Proc*. (2008) 727–731.
- [7] Home Page ICD API, (n.d.). <https://icd.who.int/icdapi> (accessed October 10, 2019).
- [8] SKOS Simple Knowledge Organization System Primer, (n.d.). <https://www.w3.org/TR/skos-primer/> (accessed October 14, 2019).
- [9] 16.14.1. Introduction, (n.d.). <http://docs.openlinksw.com/virtuoso/rdfsparqlruleintro/> (accessed October 10, 2019).

Supervised Bayesian Statistical Learning to Identify Prognostic Risk Factor Patterns from Population Data

Colin J. CROOKS^{a,1}

^a*University of Nottingham, UK*

Abstract. Current methods for building risk models assume averaged uniform effects across populations. They use weighted sums of individual risk factors from regression models with only a few interactions, such as age. This does not allow risk factor effects to vary in different morbidity contexts. This study modified a supervised Bayesian statistical learning method of topic modelling, allowing individual factors to have different effects depending on a patient's other comorbidity. This study used topic modelling to assess more than 71,000 unique risk factors in a population cohort of 1.4 million adults within routine data. The model learnt prognostically important risk factor patterns that predicted 5 year survival, and the resulting model achieved excellent calibration and discrimination with a C statistic of 0.9 in a held out validation cohort. The model explained 92\% of the observed variation in 5 year survival in the population. This paper validates using survival supervised Bayesian topic modelling within large routine electronic population health data to identify prognostically important risk factor patterns.

Keywords. Latent Dirichlet allocation, topic modelling, Bayesian modelling, electronic health records, co-morbidity

1. Introduction

To understand the effects of large numbers of individual factors, and how these vary in different combinations, requires a different approach from weighted linear summations of small numbers of individual factors common in current prediction models. Methods are needed in which risk factors are modelled by their observed patterns, rather than as individual factors. Bayesian statistical learning methods using topic modelling provide a more interpretable alternative to current machine learning methods for identifying unknown patterns in data[1] By using mixtures of risk factor patterns, risk factor effects can vary in different patients depending on the co-occurrence of other factors. This reflects the heterogeneity observed in the general population. Importantly, it has an advantage of transparency through the interpretable intermediate step of topics constructed from risk factor patterns. Topic modelling also has computational advantages through Bayesian learning, as it can process larger numbers of codes, whilst it incorporates priors to reduce over fitting where the data is sparse. Importantly, topic

¹ Corresponding Author: Colin J. Crooks, NIHR Nottingham Biomedical Research Centre at the Nottingham University Hospitals NHS Trust and the University of Nottingham, and Nottingham Digestive Diseases Centre, School of Medicine, University of Nottingham, NG7 2UH, UK; E-mail: colin.crooks@nottingham.ac.uk

modelling can be supervised by outcomes, including survival time [2, 3]. The aim of this study was to assess the feasibility of using survival supervised topic modelling to predict mortality from unselected population based routine health data through learning prognostically important morbidity patterns.

2. Methods

The study cohort consisted of all adults in the English population over 18 years who were alive and registered on 1st January 2010 to the Clinical Practice Research Datalink (CPRD, protocol number 16_269R was approved by the Independent Scientific Advisory Committee (ISAC)). All their routine primary care (diagnostic Read codes from the CPRD clinical file) and secondary care (ICD 10 diagnostic codes from linked Hospital Episodes Statistics (HES)) data recorded prior to 1st November 2009 were included in the study with no filtering. The two month exclusion window avoided including codes related to a final illness event. All deaths from this cohort were identified for the study outcome between 1st January 2010 and 1st January 2015. Each diagnostic code was weighted with the frequency with which it was recorded in the data for the year prior to 1st November 2010. Age was adjusted within the baseline hazard by setting age as the origin of follow up time. The cohort was divided randomly into a 50% training cohort, 25% testing cohort and 25% validation cohort using simple random sampling.

2.1. Bayesian statistical learning model - Survival Supervised Topic Modelling

Blei *et al* developed the initial Latent Dirichlet allocation method for categorising documents based on associating word frequency combinations with unobserved (or latent) topics[1], and observed topics (supervised allocation)[2]. For this study topic modelling was supervised by censored survival data, using cumulative risks predicted from survival models as previously described [3]. However, to scale this up to large but noisy and sparse electronic health record data, with tens of thousands of unique codes, this study integrated 2 further modifications into the algorithm: 1) An asymmetric Dirichlet prior that was learnt from the data, promoting uninformative but frequently occurring codes to be allocated into uninformative baseline topics, 2) A penalised Cox proportional hazards model fitted via a coordinate descent algorithm to improve stability, with automated centering of topics based on the current topic allocations.

2.2. An asymmetric Dirichlet Prior learnt from the data

High frequency but uninformative terms can, by their association with other terms, become allocated to informative topics. To reduce this an asymmetric prior for document level topic distributions was learnt from the word topic allocations [5]. This allowed some topics within the model to automatically form baseline topics with a higher prior Dirichlet probability, thus promoting patients with few or uninformative risk factors to be allocated to these baseline topics along with their uninformative risk factor combinations. This topic prior was given a Dirichlet distribution learnt using the Newton-Raphson method [5]. The baseline topic with the highest prevalence, i.e. the highest number of words assigned to it was allocated a zero coefficient in the Cox

model for each iteration. This aided the convergence of the Cox model by keeping the topic simplex centered.

2.3. A penalised proportional hazards model via a coordinate descent algorithm

To improve stability a penalty was introduced to the Cox model, shrinking coefficients towards the null, reducing instability, and preventing over-fitting where the data are sparse. Within the learning algorithm each topic is required to retain some level of probability, therefore L2 regularisation (or ridge regression) rather than L1 (Lasso) weighting was used [4]. Cyclic coordinate descent was used to fit the Cox model and the latent Dirichlet algorithm, as the large number of parameters was prohibitive for matrix inversion [4]. This also allowed coarse parallelisation of the document level updates.

3. Results

The topic model algorithm described above was applied to an unselected population cohort of training 712,868 patients derived from linked primary and secondary cared data from the English Clinical Practice Research Datalink for a range of numbers of topics from 10 to 200. The model was fitted to predict five year survival from patients alive on the 1st January 2010 using over 71,141 unique diagnostic codes available prior to 1st November 2009. The model converged on average in under 20 iterations. In addition to varying the topic number, a range of L2 penalties was assessed from 10^{-4} to 10^4 , with the best fitting selected by out of sample likelihood in cross validation during the maximisation step.

3.1. Model parameter selection

The model learnt in the training cohort was applied to the testing cohort of 356,482 patients, and the perplexity calculated as a measure of model fit perplexity is shown in figure 1. Perplexity is an average of the inverse likelihood per code, and the lower the perplexity the better the model fit. Figure 1 shows that model fit increased rapidly with an increase in topic number, but after 90 topics this plateaued. A fixed penalty of 0.01 was the minimal value that aided convergence in the test data and was selected as the L2 penalty. Fifty three of the topics were significantly associated with survival ($p < 0.05$ likelihood ratio test), and the remainder were null and uninformative.

The topics generated by the latent Dirichlet allocation demonstrated recognisable combinations of risk factors that cut across different disease categories to define clinically recognisable frailty and multi-morbidity types. Summaries of three of selected topics that were strongly associated with survival are shown in (table 1).

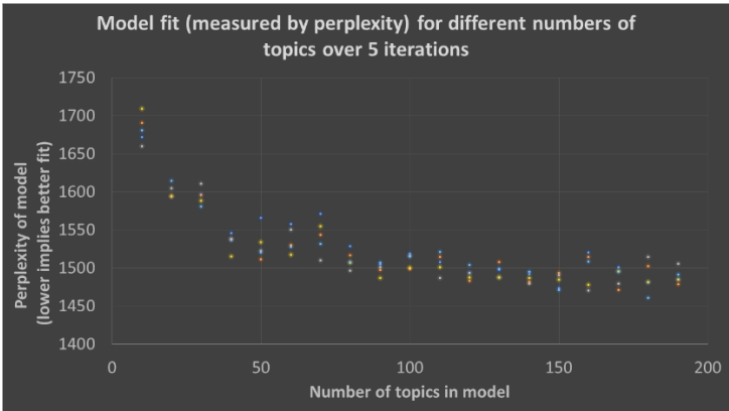


Figure 1. Model fit by number of topics in the testing dataset (lower is better)

Table 1. ICD and Read code descriptors with high probabilities for three example learnt topics that were strongly associated with survival

Summary	Cardiac disease, housebound, limited mobility and constipation	Cataract, falls, osteoporosis, fractures and UTI,	Chronic leg ulcers and nursing care
Read and ICD 10 code descriptors	• Congestive heart failure	• Hypertension	• Dressing of wound
	• Syncope and collapse	• Cataract	• Nursing care - dressing
	• AF & flutter	• High BMI	• Leg ulcer NOS
	• COPD	• Fall-accidental	• Cellulitis NOS
	• Senility	• Dementia	• Dressing of ulcer
	• Living alone	• UTI	• Wound healing
	• Constipation	• Osteoporosis	• Change of dressing
	• Housebound	• Fracture neck of femur	• Wound observation

3.2. Validation and calibration

The 90 risk factor topics that were learnt by the model in the training cohort were applied to the held out validation cohort (n=356,473 patients) using the inference algorithm as described by Blei *et al* [1], adjusted for age (in 10 year age bands) and gender. This model demonstrated a 1 year discrimination with a C statistic of 0.925 (95% confidence interval 0.921-0.929), which reduced to 0.908 (0.906-0.910) by 5 years. The average observed survival for each centile of the predicted topic model is plotted in figure 2. This showed excellent calibration across the range of predicted survival. 92% of the observed variation in 5 year survival was explained by the model (R^2 adjusted for censoring).

4. Conclusion

Topic modelling provides a novel Bayesian statistical learning approach to learn prognostic patterns of risk factors from routine health data, and this study demonstrates

it can be successfully adapted to censored electronic health records. This study confirmed the importance of using risk factor patterns to model risk rather than individual risk factors, and demonstrated how these different patterns of coding can be captured automatically. One limitation is the model assumed learnt patterns were independent from each other, and further work will incorporate a correlated topic model. However, the current method identified clinically recognisable risk factor patterns of multi-morbidity and frailty that would not be included in standard co-morbidity scores. This approach could be applied to identify cross cutting morbidity patterns that are necessary for prediction models with the increasing complexity of multi-morbidity in the population. The C++ code underlying this paper can be accessed at doi.org/10.5281/zenodo.1045521 or github.com/ColinCrooks/SurvivalSupervisedLDA.

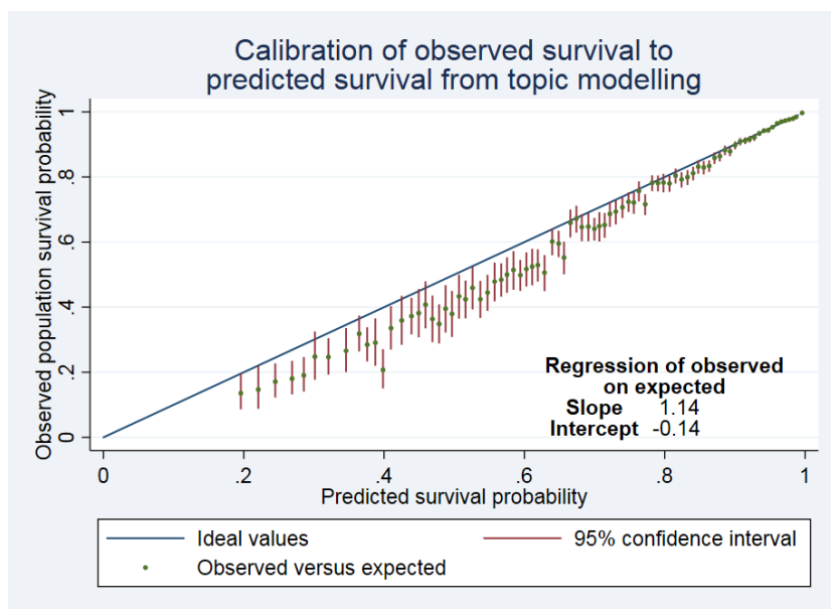


Figure 2: Calibration curve of the fitted model to the held out validation cohort

References

- [1] Blei DM, Ng A, Jordan M. Latent Dirichlet Allocation. *The Journal of Machine Learning Research*. 3 (2003), 993–1022. doi:10.5555/944919.944937.
- [2] Blei DM, McAuliffe JD. Supervised Topic Models. *Advances in Neural Information Processing Systems*. 2 (2007), 121–128.
- [3] Ye S, Dawson JA, Kendziorski C. Extending information retrieval methods to personalized genomic-based studies of disease. *Cancer informatics*. 13(Suppl 7) (2014), 85–95. doi:10.4137/CIN.S16354.
- [4] Mittal S, Madigan D, Burd RS, Suchard Ma. High-dimensional, massive sample-size Cox proportional hazards regression for survival analysis. *Biostatistics*. 15(2) (2014), 207–221. doi:10.1093/biostatistics/kxt043.
- [5] Wallach HM, Mimno D, Mccallum A. Rethinking LDA : Why Priors Matter. *Advances in Neural Information Processing Systems*. 22(2) (2009) 1973–1981. doi:10.1007/s10708-008-9161-9.

Supervised Learning for the ICD-10 Coding of French Clinical Narratives

Clément DALLOUX^a, Vincent CLAVEAU^a, Marc CUGGIA^b,
Guillaume BOUZILLÉ^b and Natalia GRABAR^c

^aUniv Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France

^bUniv Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France

^cUMR 8163 STL CNRS, Université de Lille, France

Abstract. Automatic detection of ICD-10 codes in clinical documents has become a necessity. In this article, after a brief reminder of the existing work, we present a corpus of French clinical narratives annotated with the ICD-10 codes. Then, we propose automatic methods based on neural network approaches for the automatic detection of the ICD-10 codes. The results show that we need 1) more examples per class given the number of classes to assign, and 2) a better word/concept vector representation of documents in order to accurately assign codes.

Keywords. ICD-10 coding, clinical NLP, clinical narratives, French

1. Introduction

Automatic coding of clinical narratives in French with the International Classification of Diseases, 10th revision (ICD-10) has become a necessity, as ICD-10 is used for billing, epidemiology survey and many other tasks. In this work, we present 1) a corpus with French clinical documents annotated with the ICD-10 codes, and 2) methods relying on neural networks for document-level classification. Given the difficulty of this task, the aim of our work is to create a tool that suggests relevant codes to human coders.

Several shared tasks are related to the ICD coding. The first of the kind, Computational Medicine Challenge [1] in 2007, was dedicated to the assignment of the ICD-9-CM codes to clinical narratives (radiology reports). While [2] is the first work on ICD coding in French, since 2016, CLEF eHealth offers a clinical information extraction challenge which aim is to extract ICD-10 codes from health reports. Recently [3,4,5], the CépiDC task consisted in extracting ICD-10 codes from death reports in several languages (French in 2016, French and English in 2017, French, Hungarian and Italian in 2018). In 2019 [6], the task involved the automatic annotation of German non-technical summaries of animal experiments with ICD-10 codes. The 2020 task will focus on ICD-10 coding for clinical narratives in Spanish. As we are working with French clinical documents, we next describe the 2018 French data and the most efficient systems.

The French death certificates are stored in two files: one raw text file in which the physician writes the report (raw causes) and one file containing the corresponding ICD codes (computed causes). The computed causes file might contain the text supporting

line	text	normalized text	ICD codes
1	choc septique	choc septique	A41.9
2	peritonite stercorale sur perforation colique	peritonite stercorale	K65.9
2	peritonite stercorale sur perforation colique	perforation colique	K63.1
3	Syndrome de détresse respiratoire aiguë	syndrome détresse respiratoire aiguë	J80.0

Table 1. A sample document from the CépiDC French Death Certificates Corpus: aligned dataset [5]

Aligned				Raw			
Team	Precision	Recall	F-measure	Team	Precision	Recall	F-measure
IxaMed-2	.841	.835	.838	IxaMed-1	.872	.597	.709
IxaMed-1	.846	.822	.834	IxaMed-2	.877	.588	.704
IAM-2	.794	.779	.786	LSI-UNED-1	.842	.556	.670
IAM-1	.782	.772	.777	LSI-UNED-2	.879	.540	.669

Table 2. Best CLEF eHealth 2018 results on the raw and aligned French dataset [5]

the coding decisions. Two datasets were provided for this task: one raw and one aligned. The raw dataset contains the original death certificates and the corresponding computed causes. The aligned dataset contains per line: raw text, normalized text and the corresponding ICD code, as exemplified in Table 1.

The *IxaMed* team [7] used a language-independent solution in which a neural sequence-to-sequence modeling system encodes a variable-length input sequence of tokens into a sequence of vector representations, and decodes those representations into a sequence of output tokens (the ICD10 codes). As shown in Table 2, *IxaMed* got the best results on the aligned and raw datasets. The *IAM-ISPED* team [8] used a dictionary-based approach to assign ICD-10 codes to each text line, and Levenshtein distance and synonym expansion system to detect typos. *IAM-ISPED* got the second best score on the aligned dataset and third on the raw dataset. The *LSI-UNED* team [9] submitted two runs for each dataset. Their best run used a multilayer perceptron and an One-vs-Rest (OVR) strategy supplemented with IR methods. The models were trained with the training data and dictionaries of CépiDC, estimating the frequency of terms weighted with Bi-Normal Separation (BNS). *LSI-UNED* was the second best system on the French raw dataset.

The task addressed during this challenge is interesting although the death certificates cannot be considered as free-text but rather as controlled text documents. This is especially true for the aligned dataset, which makes it very convenient for the methods proposed during the challenge. In this work, we use clinical narratives containing mostly free-text, therefore, the methods used during the challenge are not appropriate.

2. Method

Dataset. Our dataset contains 28,000 clinical documents from the Rennes university hospital. Each document contains one or several notes for a single patient issued from a single stay at the hospital. Table 3 presents statistics on our corpus. As it shows, although the tokenization methods might differ, our documents are much longer than death certificates. Indeed, the latter only contain on average 3 lines, 10 tokens and 4 ICD codes per document while our clinical narratives have on average 1,345 tokens per document and twice as many codes. It also shows that a majority of ICD codes have very few examples in the corpus which will be an issue for the deep learning methods. A typical document describes the patient’s medical history, reasons for hospitalization, laboratory measurements, findings, treatment and conclusion. Our set of data also contains some noise (e.g.

	Training	Development	Test	Total
Documents	22,400	2,800	2,800	28,000
Tokens	30,307,091	3,718,715	3,654,599	37,680,405
Total ICD codes	182,112	23,010	22,561	227,683
Unique ICD codes	5,735	2,824	2,887	6,116
Unique unseen ICD codes	-	191	212	-
Codes with less than 10 examples	3,885	2,320	2391	-
Codes with 100 examples or more	382	32	31	-

Table 3. Descriptive statistics of the Rennes University Hospital dataset

phone numbers, names, email addresses, titles, etc.) which was not preprocessed. As our data contain confidential medical information, we cannot show any examples.

Tasks. The first task addressed is the multi-label classification done on a document level for all codes. As Table 3 shows, there are 6,116 different classes (ICD codes) to identify: 5,735 codes are present in the training set among which 3,885 examples have less than 10 examples. For the second task, we reduced the hierarchical level of the ICD-10 codes. All codes with 3 digits and more were reduced to 2 digits: A01.3 *Paratyphoid fever* C becomes A01 *Typhoid and paratyphoid fevers*. Therefore, only codes A00 to Z99 are exploited, which decreases the number of classes to 1,549 overall, among which 1,501 are present in the training set.

Methods. The purpose of our work is to design an approach for the automatic and efficient detection of the ICD codes. The approaches proposed and tested rely on specifically trained word embeddings and supervised learning techniques. In the existing work, various methods have been used to represent words as vectors. Recently, several machine learning-based approaches have been introduced for a better representation of semantic relations between words. We experiment with fastText vectors [10], which use subword information to represent words missing in the training vocabulary. Each word is represented as a bag of all possible n-grams of characters it contains. Our model is trained on our dataset with the *Skip-Gram* algorithm, 300 dimensions, a window of 5 words before and after each word, a minimum count of five occurrences for each word and negative sampling. Besides, we experiment with two neural network architectures: recurrent and convolutional. A recurrent neural network is a class of network which is capable of adapting its decision by taking into account the previously seen data, in addition to the currently seen data. This operation is implemented thanks to the loops in the architecture of the network, which allows the information previously seen (previous words) to persist in memory. In our experiments, we used a bidirectional recurrent neural network with long short-term memory cells (BiLSTM). Convolutional neural networks (CNN) are widely used in Computer Vision and Natural Language Processing. The convolution layer consists of a set of filters, where each filter is convolved with the input forming feature maps. These filters are randomly initialized and become parameters that will be learned by the network.

Experiments. We trained several CNN models based on the architecture proposed in [11] for 30 epochs. We used four filter windows (2, 3, 4 or 5 width) with either 100, 500, 1,000 or 2,000 feature maps each (ReLU activation), 0.5 dropout, global max pooling and a fully connected layer with ReLU activation of either 100, 1,000 or 2,000 dimensions for a more compact representation. Regarding our BiLSTM, the dimensionality of the

output space was set to 600 units per layer (backward/forward) with 0.5 dropout. Both architectures use our pre-trained fastText vectors with 0.5 dropout and a fully connected sigmoid layer for prediction. Performance was measured with the standard metrics: precision, recall and F-measure.

3. Results

As Table 4 shows, on both tasks, our CNN, even at the lowest settings, performs better than our BiLSTM. Our experiments also show that increasing the number of feature maps per convolution filter and hidden units in the fully connected layer both improves the results. Indeed, we get an 8.21 point increase at the most expensive settings when compared to the lowest settings. We might be able to slightly improve the results if we raise parameters further. However, it would be a low increase with a great computing cost. Indeed, the results on Task 2 show a very low improvement (0.28 point) when doubling the value of the parameters.

	System	Feature maps	Hidden units	Precision	Recall	F-measure
Task 1	BiLSTM	-	600	0.6513	0.1387	0.2287
	CNN	100	100	0.6510	0.2091	0.3165
	CNN	100	1,000	0.4916	0.2491	0.3306
	CNN	500	100	0.5624	0.2652	0.3605
	CNN	500	1,000	0.4935	0.3042	0.3763
	CNN	1,000	100	0.5353	0.2597	0.3498
	CNN	1,000	1,000	0.5052	0.3200	0.3918
	CNN	2,000	2,000	0.5029	0.3301	0.3986
Task 2	BiLSTM	-	600	0.6309	0.2837	0.3914
	CNN	1,000	1,000	0.6306	0.4427	0.5202
	CNN	2,000	2,000	0.6018	0.4625	0.5230

Table 4. Results for the ICD coding tasks. The results are given in terms of Precision, Recall and F-measure. The best scores are in bold.

4. Discussion

As expected, we get better results on our second task which shows that the reduction of the number of classes is efficient. However, as we can only assign each class once to a document, we might lose some information as a document can have multiple codes belonging to the same hierarchically higher class. Besides, we get much lower results than the systems presented at CLEF eHealth 2018. This can be explained by the nature of the documents: short, structured death certificates cannot be compared to clinical narratives with free text. The tasks are therefore different. Another explanation is that some codes might not be explained by the textual data but rather by the structured data in the hospital database.

5. Conclusion

The interest for the automatic detection of ICD-10 codes has increased in recent years. Yet, the data used in previous shared tasks do not seem representative of real clinical

narratives. In our work, after a brief reminder of the existing work, we presented a new body of French clinical narratives annotated with the ICD-10 codes. Another contribution of our work is the exploitation of fastText vectors and neural networks for the automatic ICD-10 coding. The experiments are conducted and evaluated on French clinical narratives (free-text), which have not been experienced much in the previous work of this type. From a more technical point of view, our work also indicates that the convolutional neural architectures are more efficient than the recurrent neural architectures for the multi-label classification. We plan to improve our neural network performance by providing a richer feature set. Indeed, adding available structured data and using recent embedding techniques, such as BERT [12], may provide more accurate representation of the documents.

Acknowledgments. This work was partly funded by the French government support granted to the CominLabs LabEx managed by the ANR in Investing for the Future pro-gram under reference ANR-10-LABX-07-01.

References

- [1] Pestian, John P., Brew, Christopher, Matykiewicz, Pawe l, Hovermale, D. J., John-son, Neil, Cohen, K. Bretonnel, Duch, Wlodzislaw: A Shared Task Involving Multi-label Classification of Clinical Free Text. In: Proceedings of the Workshop onBioNLP 2007: Biological, Translational, and Clinical Language Processing, BioNLP'07, pp. 97–104. Association for Computational Linguistics, Stroudsburg, PA, USA2007.
- [2] Ruch, P., Gobeill, J., Tbahriti, I., & Geissbühler, A. (2008). From episodes of care to diagnosis codes: automatic text categorization for medico-economic encoding. In AMIA Annual Symposium Proceedings (Vol. 2008, p. 636). American Medical Informatics Association.
- [3] Névél, A., Cohen, K. B., Grouin, C., Hamon, T., Lavergne, T., Kelly, L., ... & Zweigenbaum, P. (2016, September). Clinical information extraction at the CLEF eHealth evaluation lab 2016. In CEUR workshop proceedings (Vol. 1609, p. 28). NIH Public Access.
- [4] Névél, A., Robert, A., Anderson, R., Cohen, K. B., Grouin, C., Lavergne, T., ... & Zweigenbaum, P. (2017, September). CLEF eHealth 2017 Multilingual Information Extraction task Overview: ICD10 Coding of Death Certificates in English and French. In CLEF (Working Notes).
- [5] Névél, A., Robert, A., Grippo, F., Morgand, C., Orsi, C., Pelikan, L., ... & Zweigenbaum, P. (2018, September). CLEF eHealth 2018 Multilingual Information Extraction Task Overview: ICD10 Coding of Death Certificates in French, Hungarian and Italian. In CLEF (Working Notes).
- [6] Kelly, L., Suominen, H., Goeuriot, L., Neves, M., Kanoulas, E., Li, D., ... & Palotti, J. (2019, September). Overview of the CLEF eHealth evaluation lab 2019. In International Conference of the Cross-Language Evaluation Forum for European Languages (pp. 322-339). Springer, Cham.
- [7] Atutxa A, Casillas A, Ezeiza N, Goenaga I, Fresno V, Gojenola K, Martinez R, Oronoz M and Perez-de-Viñaspre O (2018). IxaMed at CLEF eHealth 2018 Task 1: ICD10 Coding with a Sequence-to-Sequence approach. CLEF 2018 Online Working Notes. CEUR-WS
- [8] Cossin S, Jouhet V, Mougin F, Diallo G, and Thiessard F (2018). IAM at CLEF eHealth 2018: Concept Annotation and Coding in French Death Certificates. CLEF 2018 Online Working Notes. CEUR-WS
- [9] Almagro M, Montalvo S, Diaz de Ilarraz A, and Pérez A (2018). LSI UNED at CLEF eHealth 2018: A Combination of Information Retrieval Techniques and Neural Networks for ICD-10 Coding of Death Certificates. CLEF 2018 Online Working Notes. CEUR-WS
- [10] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135-146.
- [11] Kim, Y. (2014). Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746-1751. Association for Computational Linguistics, Doha, Qatar.
- [12] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding.

The Impact of Specialized Corpora for Word Embeddings in Natural Language Understanding

Antoine NEURAZ^{a,b}, Bastien RANCE^a, Nicolas GARCELON^a, Leonardo Campillos LLANOS^b, Anita BURGUN^a, Sophie ROSSET^b

^aINSERM, UMR 1138 Team 22, Paris Descartes, Paris, France

^bLIMSI, CNRS, Université Paris Saclay

Abstract. Recent studies in the biomedical domain suggest that learning statistical word representations (static or contextualized word embeddings) on large corpora of specialized data improve the results on downstream natural language processing (NLP) tasks. In this paper, we explore the impact of the data source of word representations on a natural language understanding task. We compared embeddings learned with Fasttext (static embedding) and ELMo (contextualized embedding) representations, learned either on the general domain (Wikipedia) or on specialized data (electronic health records, EHR). The best results were obtained with ELMo representations learned on EHR data for the two sub-tasks (+7% and +4% of gain in F1-score). Moreover, ELMo representations were trained with only a fraction of the data used for Fasttext.

Keywords. Natural Language processing, Contextual word embeddings, Natural language understanding

1. Introduction

The recent advances in language representation such as static word embeddings [1,2], and contextual word embeddings (e.g. ELMo[3]) led to a significant performances improvement in natural language understanding (NLU). These methods can take advantage of large text corpora to learn a representation of the vocabulary in the vector space according to words semantics and syntactic use[1]. These approaches are unsupervised: they do not require learning datasets with manual annotations. However, the training corpus must be large enough to learn a correct representation. For example, it is customary to use a full dump of Wikipedia® to learn the models[1–3]. In the biomedical domain, the vocabulary is very specific. Moreover, when looking at clinical reports in electronic health records (EHRs), the syntax is also different from articles from sources similar to Wikipedia. A recent study from Wang *et al.*[4] evaluated the performances of embeddings learned on Wikipedia, biomedical publications and clinical notes in English : learning embeddings using clinical notes from EHRs increased the performances. Would this improvement also be true for contextual word embeddings? Nonetheless, it may be challenging to have access to the large number of documents requested (*e.g.* more than 100k patients in the study of Wang *et al.*) or even have access to embedding matrices learned on such corpus, due to privacy issues. It is possible that using more advanced methods to learn the word representations such as

ELMo compared to static embeddings would allow researchers to use embeddings learned on general domain corpora and still obtain good performances.

The aim of this work is to explore the impact of the method and data source of embeddings in the context of a weakly supervised NLU task in French. For this task in a restricted domain (biomedicine), we focused on the two NLU tasks: slot-filling and intent classification.

More specifically, this task aims at providing NLU in a virtual assistant in French for clinicians to explore biological tests results information in the patient's record in natural language (e.g. *Donne moi le dernier résultat de créatinine* 'Give me the last result of creatinin'). The set of queries that a physician may have about characteristics and results of a patient is broad and diverse. Therefore, enabling queries in natural language may help accessing information more efficiently. Given that no public dataset is available for this task in French, we used a weakly supervised approach by generating training data from question templates, terminologies and paraphrases as described in a previous work [5].

We compared the performances of two types of word representations: static word embeddings (Fasttext) and contextualized word embeddings (ELMo). For each method, we also compared two different learning sets in French: #1 Wikipedia or #2 a set of 1M clinical notes from our local EHR.

2. Methods

2.1. Embeddings

We compared two types of embeddings: #1 continuous skip-gram model with sub-word information as implemented in fastText[2], #2 embeddings from language models (ELMo) where the vectors are learned from the internal states of a deep bidirectional language model[3]. As a baseline, we use a continuous skip-gram model learned only on the training set (no external dataset).

We also compared the performances of these methods when learned on either a general domain dataset or a specialized dataset. The general domain dataset (hereafter, Wiki) is made of a dump of the French version of Wikipedia plus the French dataset of CommonCrawl. For this dataset, we used pre-learned models for French downloaded from fasttext website¹ for Fasttext and from github² for ELMo.

The specialized dataset (hereafter, EHR) is constituted of a random set of 1M clinical notes from the clinical data warehouse of Necker – Enfants malades hospital, a French AP-HP childrens hospital in Paris[6]. This dataset contains 162M tokens and a vocabulary of size 92k. For Fasttext, we used vectors of 300 dimensions and a window-size of 5. For ELMo, we kept only a subset of the EHR data (24M tokens) due to the high training time of ELMo. To compare embeddings from different sources, we kept the hyper-parameters as described in [3].

¹ <https://fasttext.cc>

² <https://github.com/HIT-SCIR/ELMoForManyLangs>

2.2. Task

We evaluated the impact of the different embeddings approaches on the task of NLU in a virtual assistant (VA task). Given that no public dataset is available for this task in French, we used the dataset generated from templates, terminologies and paraphrases as described in a previous work [5]. The training dataset contains 16,000 questions, 144k words (mean length of a question is 9), the development set of 4,000 questions for the tuning of the models. For the evaluation, we collected from physicians in our hospital, a set of 178 questions that they would like to ask in such a system. This set of questions has been manually annotated.

The slot filling task (VA-sequence) consists of a sequence labeling task aiming at identifying, in the question, the labtest mention (e.g. *créatinine* ‘creatinin’) and the date- related information (e.g. *22/04/2012*, *depuis 4 semaines* ‘for 4 weeks’). In the training set, the number of distinct lab mentions was 336 with a length ranging from 1 to 11 tokens and a median length of 2. In terms of vocabulary, there is only an overlap of 28% between the training set and the test set for labtest mentions. The date labels include exact dates, relative dates and time ranges.[5]

The intent classification task (VA-classification) is divided into 4 sub-tasks corresponding to 4 axes of classification. For each utterance, we assign one label per axis. Two axes concern the results of the lab exams (*i.e.* the type of result (5 categories) and interpretation of the result (5 categories)). The two latter concern temporal aspects (*i.e.* the time of result (3 categories) and constraints on time (4 categories)).[5]

2.3. Models

We evaluated 5 different configurations of embeddings: continuous skip-gram model of 300 dimensions (D) learned on the training set; fasttext of 300D learned on Wiki; fasttext of 300D learned on EHR; ELMo of 1024D learned on Wiki; ELMo of 1024D learned on EHR.

For the sequence labeling task (VA-sequence), we used a recurrent neural network (RNN) based on bidirectional long short term memory units (biLSTM)[7]. We used two layers of biLSTM of size 256.

For the classification tasks we used a convolutional neural network (CNN)[8]. The model contains a 1D convolutional layer of 250 units, kernel size of 3, ReLU activation, followed by a max pooling layer and a dense fully connected layer.

All the models were implemented using Keras, with a Tensorflow backend, and the optimizer was Adam. We used dropouts after the embedding layer and before the final dense layer as well as L2-regularization on the convolutional layers to limit overfitting. We used a weighted F1-score (harmonic mean of the precision and the recall) to evaluate the results of the different models. To estimate the variability of the performances, we used 10 repetitions of 5 fold cross-validation on the test set.

3. Results

All the results are presented in Table 1. For the sequence labeling task (VA-sequence), the best results are obtained with ELMo learned on EHR with a F1-score of 0.76 (95%CI [0.74-77]) compared to Fasttext on EHR (0.67, 95%CI [0.61-0.73]). Interestingly, we

show similar results between ELMo on Wiki (0.69, 95%CI [0.67-0.70]) and Fasttext on EHR (0.67, 95%CI [0.61-0.73]). To note, the distance between the training set and the test set, estimated using the perplexity of the n-gram model, is 194.

Table 1. Results for the NLU task for the virtual assistant: sequence labeling (Bi-LSTM) and intent classification (CNN)

Method	Sequence labeling	Intent classification
	F1-score [95%CI]	Mean F1-score [95%CI]
Baseline (only training set)	0.62 [0.61-0.64]	0.62 [0.58-0.65]
Fasttext on Wiki	0.69 [0.67-0.70]	0.52 [0.49-0.55]
Fasttext on EHR	0.67 [0.61-0.73]	0.66 [0.63-0.69]
ELMo on Wiki	0.69 [0.67-0.70]	0.49 [0.46-0.52]
ELMo on EHR	0.76 [0.74-0.77]	0.70 [0.67-0.73]

For the intent classification task (VA-classification), again, the best results come from the ELMo model learned on EHR (F1-score 0.70, 95%CI [0.67-0.73]) compared to the second-best Fasttext on EHR (F1-score 0.66, 95%CI [0.63-0.69]). In this task, the models learned on Wiki performed poorly: Fasttext on Wiki with a F1-score of 0.52 (95%CI [0.49-0.55]) and ELMo on Wiki 0.49 (95%CI [0.46-0.52]) with a baseline at 0.62 (95%CI [0.58-0.65]).

4. Discussion

The use of pre-learned embeddings may improve substantially the results on certain downstream tasks. As shown by Wang *et al.* [4] on various tasks in English, we show increased performances on two of three tasks in French when using pre-learned embeddings on a large corpus of clinical notes. The higher improvement was obtained on the VA-sequence task suggesting that tasks requiring extended lexicons may benefit the most of the pre-embeddings. Indeed, the identification of labtest mentions highly depends on the vocabulary available during training and we show that 72% of the mentions in the test set are absent from the training set. In this scenario, the representations learned on a large specialized dataset helped the model to obtain a better generalization when dealing with previously unseen mentions.

The test set shows 29% of out of vocabulary words overall. It may also partly explain the improvements observed on the classification task also (VA-classification). Interestingly, in the VA-classification task, both embeddings (Fasttext and ELMo) learned on Wiki did worsen the results compared to the baseline. This noise might come from the lack of specialized vocabulary in the Wiki corpus and a different usage of specialized words that may vary depending on the biomedical context and that are not taken into account in the general domain.

Sheikhshab *et al.* [9] evaluated ELMo embeddings on named entity recognition tasks in the biomedical domain. They also showed improved results when using an in-domain corpus to train ELMo (*i.e.* articles from Pubmed). It would be interesting to evaluate ELMo embeddings learned on a corpus of biomedical article in French but collecting such a corpus may be challenging.

This study has some limits. First, the embeddings learned on EHR came from a single hospital. This may cause some bias in the results given that the semantic particularities of the different practices may have an impact on the suitability of the embeddings for the different tasks. One possible solution to tackle this issue would be to

learn embeddings from multiple sites. However, due to privacy issues, it is not possible to transfer massive amount of hospital data. Two other approaches might be used to workaround this issue and should be explored for the biomedical domain: #1 the use of federated learning to train the models without moving the data[10]; #2 learning one embedding per site and aligning the different embeddings using unsupervised techniques. Finally, it is not obvious to determine the size of the training corpus for an embedding. It is a common practice to use all the articles from Wikipedia for example. But do we need that amount of data in a specialized domain? Comparing the impact of the size of the embedding corpus on downstream task will be of great interest.

5. Conclusions

Depending on the task, embeddings learned on large corpora can have a significant impact on NLP tasks in the biomedical domain in French. Moreover, learning these embeddings on clinical notes will increase the performances compared to general domain. As it may not be feasible to access a large corpus of clinical notes, it is still profitable to use advanced methods such as ELMo learned on general domain and obtain reasonable results. When the task does not rely on a large specialized vocabulary, the impact of external embeddings might be reduced.

References

- [1] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, in: C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26*, Curran Associates, Inc., 2013: pp. 3111–3119. <http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf> (accessed March 31, 2017).
- [2] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, Enriching word vectors with subword information, *ArXiv Preprint ArXiv:1607.04606*. (2016). <https://arxiv.org/abs/1607.04606> (accessed February 14, 2017).
- [3] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, Deep Contextualized Word Representations, in: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Association for Computational Linguistics, New Orleans, Louisiana, 2018: pp. 2227–2237. doi:10.18653/v1/N18-1202.
- [4] Y. Wang, S. Liu, N. Afzal, M. Rastegar-Mojarad, L. Wang, F. Shen, P. Kingsbury, and H. Liu, A comparison of word embeddings for the biomedical natural language processing, *Journal of Biomedical Informatics*. **87** (2018) 12–20. doi:10.1016/j.jbi.2018.09.008.
- [5] A. Neuraz, A. Burgun, L.C. Llanos, and S. Rosset, Natural language understanding for task oriented dialog in the biomedical domain in a low resources context, *Machine Learning for Health (ML4H) Workshop at NeurIPS 2018*. (2018).
- [6] N. Garcelon, A. Neuraz, R. Salomon, H. Faour, V. Benoit, A. Delapalme, A. Munnich, A. Burgun, and B. Rance, A clinician friendly data warehouse oriented toward narrative reports: Dr. Warehouse, *Journal of Biomedical Informatics*. **80** (2018) 52–63. doi:10.1016/j.jbi.2018.02.019.
- [7] A. Graves, and J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, *Neural Networks*. **18** (2005) 602–610. doi:10.1016/j.neunet.2005.06.042.
- [8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*. **86** (1998) 2278–2324. doi:10.1109/5.726791.
- [9] G. Sheikhshabbafghi, I. Birol, and A. Sarkar, In-domain Context-aware Token Embeddings Improve Biomedical Named Entity Recognition, (n.d.) 5.
- [10] D. Zhang, X. Chen, D. Wang, and J. Shi, A Survey on Collaborative Deep Learning and Privacy-Preserving, in: *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, 2018: pp. 652–658. doi:10.1109/DSC.2018.00104.

The LOINC Content Model and Its Limitations of Usage in the Laboratory Domain

Cora DRENKHAHN^{a,1} and Josef INGENERF^{a,b}

^a *Institute of Medical Informatics (IMI), University of Lübeck, Germany,*

^b *IT Center for Clinical Research, Lübeck (ITCR-L), University of Lübeck, Germany*

Abstract. To unambiguously encode the semantic meaning of laboratory tests, the LOINC terminology is widely used. With regard to the constantly changing and diverse requirements of the laboratory domain, LOINC's long-established content model and related publications are reviewed conjointly, revealing some obstacles for flexible adaptation in terms of new or varying application needs as well as issues regarding the comprehensive reusability of lab data. In a concise overview, four specific limitations are identified that require adaptation or the usage of other terminologies.

Keywords. Terminologies, Semantic Interoperability, Clinical Laboratory

1. Introduction

Logical Observation Identifiers Names and Codes (LOINC) is one of the most notable and well-known terminologies in medicine and widely used for standardized coding of laboratory data in electronic health records [1]. LOINC's content model has remained unchanged since its beginnings [2] but is confronted with varying needs and requirements in modern healthcare. Here, the significance of laboratory diagnostics entails a great diversity of tasks in digitalized lab environments including order/entry of tests and reporting of their results. Distributed laboratory data often needs to be integrated as well, for example to compare a patient's test results for clinical monitoring or to aggregate and analyze data on a larger scale in a data warehouse. In all of these aspects, economic considerations are highly relevant as well, with the billing of lab procedures being part of everyday clinical practice [3]. Like all subdomains of medicine, the laboratory field is subject to constant change, but here the rapid emergence of new diagnostic tests and methods, for example in molecular genetic analysis, speeds up this process in particular [4].

In this paper, LOINC's properties are analyzed regarding the current challenges in the laboratory domain to identify present limitations of its appliance. Related recent research is incorporated to bring together individual insights for a thorough and accurate overview. This evaluation shall not be misunderstood as criticism of LOINC – its usefulness in coding lab tests is undeniable; instead we identify potential for

¹ Cora Drenkhahn, Institute of Medical Informatics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany; E-mail: c.drenkhahn@uni-luebeck.de.

improvement and cases in which another coding system such as the comprehensive ontology SNOMED CT is needed to fully cover the complex demands of laboratory medicine.

2. Material and Methods

Based on the authors’ extensive knowledge gained in previous projects, the LOINC content model and the features of its current version 2.66, as well as related literature, are critically reviewed in the context of requirements in modern laboratory diagnostics.

2.1. LOINC content model

Since its inception, LOINC is published biannually by the Regenstrief Institute as a database of unique codes associated with human-readable terms. Each code corresponds to a distinct lab test (currently 55,844 out of 91,388 codes) or another clinical observation explicitly defined by five mandatory main dimensions plus an optional method part. Therein, LOINC’s referenced properties are restricted to those required for a correct test result interpretation so that each combination represents the minimal set of values needed to distinguish this test from all others. For most axes, some further information can be included in specific subcomponents if needed for a full definition.

Component describes the kind of substance or analyte measured stating the test’s most basic information with an immense variety of possible values (currently 24,707 part codes) and complexity of contents (up to three subcomponents).

Property denotes in which characteristic or type of property the component was measured, e.g. as number, as substance concentration or the qualitative presence. Obviously, this value is highly correlated with the test result’s unit which is *not specified* in the LOINC term [5]. With 135 different options, the property axis is known to cause misunderstandings and therefore issues while mapping to LOINC [6].

Time is used to differentiate measurements at a single point of time from those over a (specified) interval of time with an optional subcomponent.

System defines the specimen or material sample used for testing as the second most distinct characteristic. Further refinement is typically needed by other properties (e.g. 32 LOINC codes for “creatinine” in “urine”) or a subcomponent.

Scale broadly differentiates between quantitative, ordinal and nominal measurements, forming an interrelation to the *Property* part. Answer options for qualitative tests shall be coded with SNOMED CT; for quantitative results exemplary value ranges and units are partially mentioned but not explicitly defined.

Method information is only included if the measurement technique is clinically significant by affecting the test’s result or reference range. So, most terms (63.32%) lack this dimension.

Table 1. Some exemplary LOINC codes with their defining characteristics of six dimensions. Subcomponents are divided by the ^ sign.

LOINC code	Component	Property	Time	System	Scale	Method
14682-9	Creatinine	SCnc	Pt	Ser/Plas	Qn	
14684-5	Creatinine	SRat	24H	Urine	Qn	
5802-1	Nitrite	PrThr	Pt	Urine	Ord	Test strip

14754-6	Glucose^1H post 50 g glucose PO	SCnc	Pt	Ser/Plas	Qn
14578-9	AB0 group	Type	Pt	Bld^BPU	Nom

Structure and content of terms and subcomponents as shown in Table 1 reveal LOINC's origin to work conjunctively with the syntactic interoperability standard HL7 v2. Omitted information like result value, unit, reference range or context are meant to be communicated in other fields of the HL7 v2 message or the (newer) HL7 FHIR Observation resource. Another design principle states to only add sensible combinations required in practice [2,7], limiting availability to existing pre-coordinated terms.

2.2. Related research

In a previous project, we investigated the mapping of laboratory services contained in an internal catalog of a German hospital to a standardized terminology. Internal catalogs capture the local range of services and are thus used for billing purposes. This leads to a primary focus on methods and environmental information, whereas measurement parameters like specimen or property are seldom mentioned. Because of this, we found it impossible to map the heterogenous contents to LOINC due to the small overlap of included information and the prerequisite to use complete pre-coordinated terms. As a solution, SNOMED CT could be utilized based on its concepts of diverse granularity and ability for post-coordination [8].

In a literature review, we found 929 publications since 2017 mentioning the terms 'LOINC', 'laboratory' plus 'limitations' or 'challenges' and further investigated 25 of these classified as most relevant by Google Scholar. Two articles covered the most relevant aspects for our review. Bietenbeck et al. assessed three terminologies including LOINC and SNOMED CT for their ability to encode laboratory results of different complexity and correlated evaluation parameters. Whereas LOINC was found to be easily usable, interpretive comments like measurement uncertainty or reference intervals could not be expressed adequately. With SNOMED CT more content could be encoded although full coverage remained elusive. Another issue concerning LOINC was detected in the non-existent formal definitions – and explicit hierarchies in particular – resulting in unclear (subclass) relations between terms and limited possibilities of computational analysis [9]. Stram et al. conducted a comprehensive evaluation of current challenges in LOINC usage in pathology laboratories. Among other results, an increased risk for mismatches between laboratories due to inconsistent mapping, especially of method properties, was revealed. Additionally, molecular genetic diagnostics were identified to pose serious problems on the current content model regarding the vast quantities of singular tests with heterogeneous features involved, e.g. for 22,000 genes. Rapidly evolving subdomains like this were also mentioned to be challenging for keeping LOINC up to date [10].

3. Results

Putting LOINC content model characteristics into the context of current challenges in lab diagnostics reveals four types of limitations and their underlying causes:

3.1. *Billing*

Although the billing of laboratory procedures is highly correlated with the creation of test results and thus combined evaluations may be of interest, both applications require different sets of attributes and can not easily be integrated with the same terminology as shown in [8]. These results can be transferred to the general scope because LOINC's usage-specific content model offers no means to reasonably represent billing data, primarily because of its limitation to precoordinated combinations. The enforced usage of five axes that are partially irrelevant for economic purposes whereas intentionally omitting billing-relevant methods lead to an unsuitable data model.

3.2. *Context and interpretive comments*

By examining the historic origin as an addition to HL7 standards and the coverage of its accordingly developed content model, LOINC turns out to be not intended as a standalone format to fully describe laboratory test results. Additional information referring to test environment, device characteristics or result specific comments are meant to be communicated with relation to but outside of the term. So, the attempt to interpret laboratory data exclusively based on LOINC code and result value is inherently futile.

3.3. *Aggregation of test results*

Test results sharing the same LOINC code cannot ensure a simple and correct integration for combined analysis due to some distorting factors. As explained in section 3.2, LOINC terms are lacking some crucial information for complete result interpretation, e.g. different devices and reference ranges leading to incomparable results. Furthermore, seemingly identical tests specified by the same LOINC code may actually refer to different observations caused by an inaccurate mapping of idiosyncratic terms in the first place. Here, the (necessary) complexity of the content model with subcomponents and obscure *Property* values can be identified as a source of error. During mapping, another issue arises from non-existent formal definitions and explicit hierarchies so that easily locating related terms of different granularity (e.g. with a specified method) is prevented. The missing formal conceptualization further restricts advanced aggregation as well. Otherwise, LOINC terms differing only in the *Property* axis may be principally evaluated together when their units of measurement are convertible into one another but both unit defining relations and conversion support are currently missing.

3.4. *Molecular genetics and other rapidly changing domains*

Because of the limitation to reviewed pre-coordinated terms, LOINC can hardly keep up with newly developing tests and methods. Restricting contents to sensible combinations has obvious advantages in terms of data quality and error prevention but sacrifices any flexibility on the other hand. This rigid content model does not scale well for the numerous and diverse tests required for genetic diagnostics.

4. Discussion

In a complex field of application like the laboratory domain, the presented list of limitations is obviously non-exhaustive, instead we focused on the apparently most urgent issues. An application of LOINC for billing purposes is clearly out of its predefined scope but has previously been discussed and is not far-fetched. For most of the above-mentioned challenges (billing, context, and aggregation), SNOMED CT provides an approach for improvement based on its larger coverage and ontological features, but no complete solution either [8,9]. A combined usage of both terminologies is explicitly favorable. Of course, improvements to LOINC are continually proposed and implemented as well, e.g. we recently developed an ontological representation including LOINC's implicit hierarchies [11], Hauser et al. introduced an automated process for unit conversion of compatible codes [12], and the HL7 Clinical Genomics Work Group published an implementation guide for genetic test reporting with LOINC [13].

5. Conclusion

The long-established content model of LOINC partially restricts the coding system's usefulness for derived, originally not-intended applications needs, for changing demands as well as for test result interpretation and aggregation. This is due to its inflexibility, lacking formalization and incomplete information. LOINC is no all-purpose, standalone format but can benefit largely from further specification, extension and combined usage with other terminologies, primarily SNOMED CT.

References

- [1] O. Bodenreider, R. Cornet, D.J. Vreeman, Recent developments in clinical terminologies - SNOMED CT, LOINC, and RxNorm, *Yearbook of medical informatics*. 27.01 (2018), 129-139.
- [2] A.W. Forrey, C.J. McDonald, G. DeMoor, S.M. Huff., D. Leavelle, D. Leland, T. Fiers, L. Charles, B. Griffin, F. Stalling, A. Tullis, Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results, *Clinical chemistry*. 42.1 (1996), 81-90.
- [3] K.D. Pagana, T.J. Pagana, *Mosby's Manual of Diagnostic and Laboratory Tests*, Elsevier Health Sciences, St. Louis, 2017.
- [4] S. Roy, W.A. LaFramboise, Y.E. Nikiforov, M.N. Nikiforova, M.J. Routbort, J. Pfeifer, R. Nagarajan, A.B. Carter; L. Pantanowitz, Next-generation sequencing informatics: challenges and strategies for implementation in a clinical environment, *Archives of pathology & laboratory medicine*. 140.9 (2016), 958-975.
- [5] D.J. Vreeman, J. Hook, B.E. Dixon, Learning from the crowd while mapping to LOINC. *Journal of the American Medical Informatics Association*. 22.6 (2015), 1205-1211.
- [6] D.J. Vreeman, Top 10 Tips for Mapping to LOINC, (2016). <https://danielvreeman.com/top-10-tips-for-mapping-to-loinc/> (accessed September 24, 2019).
- [7] C. McDonald, S. Huff, J. Deckard, S. Armson, S. Abhyankar, D.J. Vreeman, *LOINC Users' Guide*. Regenstrief Institute, Indianapolis, 2019.
- [8] C.M. Drenkhahn, H. Ulrich, P. Duhm-Harbeck, J. Ingenerf, Mapping of Laboratory Services included in an Internal Catalog to SNOMED CT using the UMLS, *Student Conference 2018, Medical Engineering Science, Medical Informatics and Biomedical Engineering*. 2018, 337-340.
- [9] A. Bietenbeck, M. Boeker, S. Schulz, NPU, LOINC, and SNOMED CT: a comparison of terminologies for laboratory results reveals individual advantages and a lack of possibilities to encode interpretive comments, *LaboratoriumsMedizin*. 42.6 (2018), 267-275.

- [10] M. Stram, T. Gigliotti, D. Hartman, A. Pitkus, S.M. Huff, M. Riben, W.H. Henricks, N. Farahani, L. Pantanowitz, Logical Observation Identifiers Names and Codes for Laboratorians: Potential Solutions and Challenges for Interoperability, *Archives of pathology & laboratory medicine*. 2019.
- [11] C. Drenkhahn, P. Duhm-Harbeck, J. Ingenerf, Aggregation and Visualization of Laboratory Data by Using Ontological Tools Based on LOINC and SNOMED CT, *Studies in health technology and informatics*, 264 (2019), 108-112.
- [12] R.G. Hauser, D.B. Quine, A. Ryder, S.Campbell, Unit conversions between LOINC codes, *Journal of the American Medical Informatics Association*. 25.2 (2017), 192-196.
- [13] J. Deckard, C.J. McDonald, D.J. Vreeman, Supporting interoperability of genetic data with LOINC. *Journal of the American Medical Informatics Association*. 22.3 (2015), 621-627.

The openEHR Genomics Project

To the memory of Gianluigi Zanetti

Cecilia MASCIA ^{a,1}, Francesca FREXIA ^a, Paolo UVA ^a, Gianluigi ZANETTI ^a,
Luca PIREDDU ^a, Gideon GIACOMELLI ^b, Christina JAEGER-SCHMIDT ^b,
Aurelie TOMCZAK ^{b,c}, Simon SCHUMACHER ^b, Florian KRAECHER ^b,
Roland EILS ^b, Silje LJOSLAND BAKKE ^d, Heather LESLIE ^d

^aCRS4: Center for Advanced Studies, Research and Development in Sardinia

^bHiGHmed Consortium

^cInstitute of Pathology, University Hospital Heidelberg

^dClinical Model program, openEHR Foundation

Abstract. Current high-throughput sequencing technologies allow us to acquire entire genomes in a very short time and at a relatively sustainable cost, thus resulting in an increasing diffusion of genetic test capabilities, in specialized clinical laboratories and research centers. In contrast, it is still limited the impact of genomic information on clinical decisions, as an effective interpretation is a challenging task. From the technological point of view, genomic data are big in size, have a complex granular nature and strongly depend on the computational steps of the generation and processing workflows. This article introduces our work to create the openEHR Genomic Project and the set of genomic information models we developed to catch such complex structure and to preserve data provenance efficiently in a machine-readable format. The models support clinical actionability of data, by improving their quality, fostering interoperability and laying the basis for re-usability.

Keywords. genomic models, openEHR, mutations, variations, structured data

1. Introduction

Within the healthcare domain, Next Generation Sequencing (NGS) data sets are a potential asset of clinically valuable information that, if properly read and deciphered, can pave the way for life-enhancing applications such as personalised therapeutic treatments or early detection of certain hereditary diseases, just to name a few. However, handling genomic data is a very complex task from both clinical and technological perspectives. Several initiatives are addressing the technological challenges [1,2], but they are mainly focused on the exchange aspects, while many facets of the subject still remain unaddressed. In the first place, genomic information is composed of many levels, derived from each calculation step, but usually the intermediate results are expressed in a wide variety of text file formats (e.g., SAM/BAM, VCF, etc.), thus hindering the efficient re-use, sharing and possible inclusion of data in an electronic health record (EHR). In addi-

¹Corresponding Author: Cecilia Mascia, C/O CRS4, Loc. Piscina Manna - Edificio 1, 09050 Pula (CA), Italia; E-mail: cecilia.mascia@crs4.it.

tion, bioinformatic workflows typically rely on many external resources (e.g., reference genome sequences, biological annotation databases, analysis tools) that evolve rapidly and have a significant impact on the final results. Hence, it is essential to capture the details of each workflow execution to ensure, for instance, accurate data auditability and reproducibility or to make results comparable when a different pipeline is used.

The availability of NGS data in a structured and machine-readable format would therefore facilitate their integration into effective clinical actions, thanks to the possibility of capturing granularity and provenance. This need has motivated the creation of genomic models in the form of openEHR archetypes [3,4], fostering the foundation of the Genomics openEHR modelling group – consisting of members from the CRS4 (Italy), the HiGHmed Consortium (Germany), the EUCANCan project (EU, Canada), the CINTESIS group (Portugal), the BigMed project (Norway) and the international openEHR Clinical Modelling Program [5]. This article describes how our collaborative work led to the launch of the openEHR Genomics Project, the results obtained so far and the future directions.

2. Method

2.1. The OpenEHR Approach and the OpenEHR Clinical Modelling Program

OpenEHR is set of open specifications for structured data, designed to represent medical knowledge separating domain semantics from any specific technology [6]. OpenEHR archetypes are the maximum set of attributes describing a clinical concept in a machine-readable format. They enable to represent concepts of different scale of granularity, to add specific constraints on elements and to embed their semantics. Conceptual models are formalised in computable entities through the Archetype Definition Language (ADL), which can be easily converted to a number of other popular formats like XML. The OpenEHR framework also includes a mechanism to aggregate archetypes in use case-oriented structures (i.e., templates), a query language optimised for archetypes (i.e., AQL-Archetype Query Language), a set of open APIs and a conformance test approach to validate real implementations.

The openEHR Clinical Modelling Program gathers clinicians, researchers and implementers in an international collaborative action for the development of multilingual archetypes and templates. The direct involvement of domain experts in the design process is one of the key strategies adopted to obtain high quality models, whose re-use is encouraged. Archetypes are collaboratively reviewed and made available to the community through the international Clinical Knowledge Manager (CKM) [7] – the openEHR artefacts repository – which has also some national instances developed to meet specific demands.

2.2. The Initial Genomic OpenEHR Archetypes Collection

The starting point was the set of 11 archetypes proposed in [3] that should be used within the *Laboratory test result* archetype to represent genetic test results, developed on the basis of the Variant Calling Format (VCF) file specification [8]. VCF files consist of a set of rows each describing a variation that occurred in a sample sequence together with

customisable functional annotations like, for instance, '*transcript features*' and the '*predicted impact*'. Within the archetypes, mutations are described according to the Sequence Variant Nomenclature² of the Human Genome Variation Society (HGVS) [9]. Further, the models included elements like the '*conservation score*' and the '*allele frequency*' that are usually not included in the medical report itself, but are potentially relevant on the bioinformatics research side. Finally, each tool or external database used in the analysis can be tracked by explicitly recording their name, version and, where appropriate, URL, thus integrating provenance within the data.

The Design Process

Two use cases drove the creation of a common semantic data model for the representation of variant information derived from whole genome and exome sequencing (WGS, WES) data:

- Rare Diseases: to identify pathogenic variants responsible for rare diseases through WES tests of patients and their parents.
- Oncology: to enable an evidence-based and better informed diagnostic and treatment design, based on molecular stratification.

The whole design process has been guided by the co-chairs of the openEHR Clinical Modelling Editorial Group (CMEG), responsible for the modeling governance within the community.

The first step was the rethinking of content and structure, when the whole set of nodes constituting the archetypes have been assessed from two opposing perspectives. Starting from the overall picture, we prepared a mind map (using the collaborative tool XMind³) to depict all the necessary pieces of information describing the use cases, to detect possible repetitions or overlapping with existent archetypes and to simplify the structure. Starting from the single nodes, we checked the adequacy of the concept name, its adherence with standard nomenclatures and the consistence with the element's description, adding examples to clarify the intended usage.

After reaching a good level of agreement, working with the Archetype Designer⁴, we converted the mind maps into archetypes, taking further decisions on data types, node occurrences and constraints and adding, when possible, bounds to external standard ontologies.

3. Results

On April 2019, the Genomics Project⁵ was officially created in the openEHR CKM. The project collects and makes publicly accessible the 13 archetypes developed in this work.

The *Genetic variant* is the core archetype designed to be nested within the *Laboratory test result* to report observations and annotations related to mutations found in the genome as the result of a sequencing test. The morphological description of the muta-

²Guidelines available at: <http://varnomen.hgvs.org/>

³XMind 8: <https://www.xmind.net/xmind8-pro/>

⁴Archetype Designer: <https://ehrscape.marand.si/designerv2/#/>

⁵CKM Genomic project: <https://ckm.openehr.org/ckm/#showProject=1013.30.50>

tions – i.e., nucleotide changes and genomic coordinates – is given at the DNA level with reference to a genomic reference sequence. The model supports all the simple change types identified in the HGVS guidelines [9] plus translocations and copy number variations – two structural variations that are often relevant in cancer studies. In addition, further descriptions of the mutations are possible at the transcript level in terms of changes in the coding sequence ('DNA changes') and in the protein ('amino acid changes'). The reference sequence used to annotate the variations is explicitly recorded too, through a dedicated archetype (*Reference sequence*). The archetype design allows the recording of annotations for each transcript affected by that specific mutation, pointing out the one with the highest predicted impact, considered relevant for the clinical evaluation. In this regards, the functional interpretation of the variants can be specified following the ACMG classification [10] or by citing relevant scientific literature. All the external resources used by the data analysis pipeline, including biological databases and software, can be explicitly recorded by giving their IDs, version and, possibly, a unique resource locator (URL); an archetype is dedicated to capturing this information in a structured manner (*Knowledge base*). In order to clearly state the concepts' semantic and improve data quality and interoperability, most of the model's nodes are codified using LOINC⁶, an international standard ontology for identifying health measurements, observations, and documents.

OpenEHR's flexibility allows to project the archetype data onto a variety of different use cases, enabling to archive all the elements automatically captured during pipeline execution, but presenting to the final user only those relevant to the task at hand. To demonstrate how the archetypes can be combined and integrated, an example template for a generic sequencing test report has been included in the Project.

The initial review round of the *Genetic variant* archetype started on April 26 2019, and, afterwards, we commenced in staggered steps the review process of each archetype, inviting all the interested domain experts in the project to give their opinions and comments. The process is managed by the project editors and performed completely via the CKM, which also allows the tracking of archetypes changes. At the moment of writing, one archetype has been published after one review round while the others are still in the consolidation phase. In the meantime, the Norwegian translations of the archetypes are currently being refined.

4. Discussion

The focus of our work was the creation of information models to capture genomic data for clinical and research applications. The openEHR community has overall expressed a good consensus on the structure of the models, together with a series of useful feedbacks. For instance, a series of suggestions has been provided to improve the coverage of the genomic domain, particularly with respect to the representation of complex mutations, such as splicing variants affecting RNA sequences, which are not yet supported by model. The review process has also provided important hints on how to improve and extend the usability of the models in real contexts, as a set of suggestions emerged to refine the requirements and to generalise the models towards different use cases than those

⁶Logical Observation Identifiers Names and Codes (LOINC): <https://loinc.org/>

initially considered. The feedback will be integrated in the next versions of the models while trying to balance the different perspectives.

More generally, the number of reviewers involved in this specific CKM domain tend to be fewer than those involved in more long-standing ones. We believe this is mainly due to the novelty and the specialised nature of the project, and that the situation can be improved with targeted actions to disseminate the potential of the obtained results.

5. Conclusion and Future Work

In this work we presented the openEHR Genomics Project and its first results: the models to represent clinical genomic data and its provenance with a set of openEHR archetypes. These results have been developed through a collaborative process supported by the openEHR Clinical Modelling Program. Capturing the complexity of genetic information remains a challenging task, but the results presented in this work are the promising starting point for an ongoing development and improvement process. Our next steps include to extend the models to cover additional concepts, like the complex genetic variants, and to continue the ongoing review process in the international CKM. Further, additional model reviews at the Norwegian and German repository instances are already planned, and more will be organized in due time.

6. Acknowledgements

The authors from CRS4 dedicate this work to the memory of Gianluigi Zanetti, whose legacy continues to inspire. This work has been partially supported by the European Joint Programme on Rare Disease (grant agreement N. 825575).

References

- [1] Genomics - FHIR v4.0.0, <https://www.hl7.org/fhir/genomics.html>, (Accessed on 09/29/2019).
- [2] Phenopackets - Concepts and Technology, <http://phenopackets.org/>, (Accessed on 09/29/2019).
- [3] C. Mascia, et al., OpenEHR modeling for genomics in clinical practice, *International Journal of Medical Informatics* 120 (2018) 147–156 (Dec. 2018). doi:10.1016/j.ijmedinf.2018.10.007.
- [4] P. A. Maranhão, et al., Challenges in design and creation of genetic openEHR-archetype, *Studies in Health Technology and Informatics* 247 (2015) (2018) 835–839 (2018). doi:10.3233/978-1-61499-852-5-835.
- [5] Genomics information - openEHR Clinical - openEHR wiki, <https://openehr.atlassian.net/wiki/spaces/healthmod/pages/50987020/Genomics+information>, (Accessed on 10/03/2019).
- [6] What is openEHR?, https://www.openehr.org/about/what_is_openehr, (Accessed on 09/10/2019).
- [7] Clinical knowledge manager, <https://ckm.openehr.org/ckm/>, (Accessed on 09/10/2019).
- [8] The Variant Call Format (VCF) Version 4.2 Specification, <https://samtools.github.io/hts-specs/VCFv4.2.pdf>, (Accessed on 09/27/2019) (2019).
- [9] J. T. Den Dunnen, et al., HGVS Recommendations for the Description of Sequence Variants: 2016 Update (2016). doi:10.1002/humu.22981.
- [10] S. Richards, et.al, Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology, *Genetics in Medicine* 17 (5) (2015) 405–423 (2015). arXiv:15334406, doi:10.1038/gim.2015.30.

The Personal Health Library: A Single Point of Secure Access to Patient Digital Health Information

Nariman AMMAR^a, James E. BAILEY^b, Robert L. DAVIS^a, and
Arash SHABAN-NEJAD^{a,1}

^a*Department of Pediatrics, College of Medicine, The University of Tennessee Health Science Center – Oak-Ridge National Laboratory (UTHSC-ORNL), Center for Bio-medical Informatics, Memphis, Tennessee, USA*

^b*Center for Health System Improvement, College of Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, USA*

Abstract. Traditionally, health data management has been EMR-based and mostly handled by health care providers. Mechanisms are needed to give patients more control over their health conditions. Personal Health Libraries (PHLs) provide a single point of secure access to patients' digital health information that can help empower patients to make better-informed decisions about their health care. This paper reports a work-in-progress on leveraging tools and methods from artificial intelligence and knowledge representation to build a private, decentralized PHL that supports interoperability and, ultimately, true care integration. We demonstrate how a social application querying such a decentralized PHL can deliver a tailored push notification intervention focused on improving self-care behaviors in diabetic adults from medically underserved communities.

Keywords. Personal health library, digital health, privacy, interoperability.

1. Introduction

Currently, patients receive important digital health information in multiple settings and through different communication modalities. Emerging evidence indicates that patients are frustrated by the lack of EMR interoperability among these fragmented systems and platforms [3-8] and want to have an active role in managing their data [1, 2]. Improved interoperability and support for patient-provider communication has the potential to improve patient satisfaction and, evidence suggests, could even help detect and prevent medical errors [8]. Table 1 summarizes some critical requirements of PHL from the patient perspective identified in the literature [1-8]. In digital health management, a health state is a digital representation of the state of someone's health at a point in time, including their medications, test results, exercises, diets, insurance claims, vaccines, allergies, appointments, and their outcomes. Such data come from diverse sources, including health care organizations, government agencies, and clinicians. It can also include digital health information from mass media and social networks. Health data also comes in different formats, including EMRs, family histories and genealogies, data streams from activity trackers, personal genome sequences, articles, videos, and public datasets. Health states change over time as data is acquired through ongoing health-

¹ Corresponding Author, E-mail: ashabann@uthsc.edu.

related processes and events embedded within those processes. A health-related intervention is a process, which could either be therapeutic or preventive. For instance, an intervention for self-management of chronic diseases (e.g. diabetes) focused on promoting health behavior change to improve clinical outcomes is an example of a preventive intervention. In this paper, we describe a strategic implementation plan for a social application querying a decentralized PHL that can deliver a tailored push notification intervention focused on improving self-care behaviors in diabetic adults from under-served communities.

Table 1. Patient requirements (Req.) for a PHL.

Req.	Description
<i>Construction and Management (PODs)</i>	
R1.1 R1.2 R1.3	A mechanism to construct a PHL by bringing their data together in a trustworthy, usable, and useful library by gathering different types of knowledge into a single resource.
R2	Ability to manage the PHL by adding, editing, or removing resources
<i>Access Management and Privacy (PODs)</i>	
R3.1 R3.2	Ability to decide what types of data should be kept, and who has access to that data
<i>Knowledge Acquisition and Aggregation</i>	
R4.1 R4.2	Ability to seek health information from constantly changing public sources, enriched with new streams and types of data
R5	Ability to decide what types of data is important enough to collect, manage, and share
R6	Ability to participate by sharing their data with citizen science and research initiatives
<i>Usage (Applications)</i>	
R7.1 R7.2	Ability to search through the PHL using intelligent mapping for vocabulary used to describe resources in their profile.
R8	Ability to annotate their own results from participating in clinical trials to look for patterns
R9	Ability to receive alerts about new information related to topics covered in the PHL
R10	Ability to play an active role in staying healthy by monitoring their progress
R12	Ability to find and use information including text summarization, knowledge mapping.
R13	Ability to get digital assistance: personalized alerts, literacy aids, translations.

2. Method

We provide a solution for constructing and using a Personal Health Library (PHL) inspired by the Social Linked Data (Solid) framework [9]. Solid builds on the future vision of the decentralized Web by providing an infrastructure for separating data from applications that process that data. It provides users with true ownership of their data by storing it among several Personal Online Data stores (PODs) hosted wherever they

desire, and selectively authenticating applications to access and process specific resources within those PODs. The proposed Solid-based PHL manages distributed data storage in several PODs and enables interoperability between applications by reusing standard protocols, formats, and vocabularies, and by providing a common access RESTful API. Table 2 shows the main features supported by the proposed Solid-based framework.

Table 2. Features supported by the proposed Solid-based framework.

F _n	Description
F1	<i>Agents</i> : Solid defines standard vocabulary to describe the different entities that contribute to a user’s profile as Agents of different types (Person, Organization, Device, etc.)
F2	<i>WebID</i> : Each agent has a unique WebID, which links to a public RDF-based profile document that may link to other extended profiles. Extended profiles represent different user’s application profiles.
F3	<i>Web of Trust</i> : Extended profiles can be used to build a Web of trust using the FOAF vocabulary by allowing agents to link their profiles in a public or protected way.
F4	<i>Web of Resources</i> : Whether it is a person, an inbox, a document, an image, a notification, or a relationship, everything within a profile is represented as a resource.
F5	<i>Hierarchical resource representation</i> : Resources can be organized hierarchically in containers (directories), which allows for fine-grained access control lists (ACL). Solid uses the WAC Ontology.
F6	<i>Resource Management</i> : RDF, the model driving the Linked Data Platform (LDP), follows REST principles of identifying resources by URIs. Users manage resources via HTTP operations on their URIs. SPARQL endpoints enable complex multi-pod data retrieval via link-following SPARQL.
F7	<i>Linkability</i> : Resources stored in different PODs are represented following LDP recommendations. LDP enables agents to dereference resource URLs to reach a resource from another. Links within LDP are typed, which enables agents to explicitly state how their linked resources are related.
F8	<i>Live Notifications</i> : are implemented following the W3C LDN protocol and <i>ActivityStream Vocabulary</i> and enable users to interact with content on each other’s PODs and enable content owners to receive notifications in the appropriate <i>inbox</i> .
F9	<i>Portability and Interoperability</i> : The framework has a standard API that makes it easy for developers to write applications that allow users to use the same data in different applications instead of locking it inside different application data repositories
F10	<i>Security: Authentication</i> : The framework provides decentralized authentication, a global id space, and a global single sign-on through WebIDs. User has to register with an identity provider (their POD provider). The provider stores the users’ WebID profile document associated with a cryptographic key.
F11	<i>Privacy</i> : Access Control Lists (ACL). Solid uses the WAC Ontology to define read, write, edit, etc. permissions on resources. Since resources are hierarchical this allows for fine-grained access control. A patient can grant fine-grained access permissions to the different agents, ranging from a single person or a group to everyone. The patient can also have a list of trusted applications and control their access to the different resources in his/her POD.

3. Results

To showcase the proposed Solid-based PHL in action, we focus on one application scenario, featuring a preventive intervention from the chronic disease self-management domain:

Bob is an African American adult with diabetes equipped with a smart-phone that collects physiological data (e.g. step counts) in real-time. The social application on his smartphone queries his decentralized PHL to deliver tailored push notifications to support behavior change related to chronic disease self-care. Depending on sensor readings and other information in his PHL, the app provides personalized and tailored recommendations for healthy eating, physical activity, medication taking, and/or visiting healthcare providers. Alice is a cancer patient who is part of Bob's social network and Mary is his primary care physician.

We leverage the Solid infrastructure and the relevant linked data principles to assist Bob in constructing and managing his PHL as well as interacting with other collaborators in his health and social networks, through the Solid ecosystem.

First, Bob generates unique WebID to securely login to his profile document (Table 2, F2, F10), sets up his PODs, and keeps a list of vCard URIs of trusted (Table2, F3) agents and applications in his extended profile. He adds Alice and Mary as Person agents and he adds the monitoring app and his smartphone as *device* and *software* agents, respectively (Table2, F1). Bob is interested in *gathering* the latest blog posts about diabetes from a blogging app. He adds the app under the trusted apps section of his profile document. He grants each of the above agents a fine-grained access permission to the corresponding containers under his POD with the proper access mode (Table2, F11). He can set up a chatting *channel* of type LongChat about Diabetes as a resource under the Diabetes folder under his public folder (Table2, F4,F5,F6).

Alice and Mary can *subscribe* to the channel using their WebIDs and have message *notifications* pushed to their inboxes as resources with unique URIs and they can *collaborate* on assessing each other's messages. Bob can also *share* blog posts that he receives from the blogging app to Alice's inbox. Alice gets a notification in her inbox of the blog post (Table2, F8). She has several options to *interact* with the document, including replying to Bob, adding *annotations* to robustify links in the document by linking them to other concepts or sources of knowledge, and save either a copy of that document to her POD or a link to Bob's copy.

Resources generated by each of the three users is stored in their corresponding PODs, but can be linked to each other (Table2, F7). For example, message notifications get pushed by the messaging POD to the inboxes of the users who subscribed to the messaging app as resources with unique URIs. Assuming a message stored in Bob's profile is identified by <https://bob.solid/messages/diabetes/1234>, then Alice's comment at <https://alice.solid/comments/36756> links back to Bob's message by the *hasTarget* Link Type defined in the Web Annotation Ontology (www.w3.org/ns/oa).

4. Discussion and Future Work

This paper has shown how a social application querying such a decentralized PHL can securely access, exchange, and use PHLs and deliver personalized and tailored push notifications to support chronic disease self-care. In this paper we employed a Solid-

based framework to assist users in constructing their PHL by gathering different types of data, information, and knowledge into a single searchable resource. While there has been some effort to build similar systems, the novelty of the proposed approach lies in i) providing a decentralized, yet linked architecture; ii) supporting interoperability, portability, knowledge mapping, and reasoning by following protocol, format, and vocabulary standards; iii) building trust with patients by facilitating true ownership over their data, and appropriate reporting; and (4) giving them a fine-grained access control mechanisms.

Future work will focus on further implementation of an end-to-end framework of intelligent recommendation and digital assistance tools, including text summarization, knowledge mapping, and personalized resource suggestions. In achieving this goal, we will incorporate AI techniques and knowledge representation methods that have been successfully used in our previous works [10, 11, 12]. Other ongoing tasks will include setting up a clinical trial of the social application and recruiting participants to fully evaluate the application.

References

- [1] Baird, A., North, F. and Raghu, T.S., 2011, February. Personal Health Records (PHR) and the future of the physician-patient relationship. In *Proceedings of the 2011 iConference* (pp. 281-288). ACM.
- [2] Laine, C. and Davidoff, F., 1996. Patient-centered medicine: a professional evolution. *Jama*, 275(2), pp. 152-156.
- [3] Auerbach, S.M., 2000. Should patients have control over their own health care?: Empirical evidence and research issues. *Annals of Behavioral Medicine*, 22(3), pp.246-259.
- [4] Ross, S.E., Todd, J., Moore, L.A., Beaty, B.L., Wittevrongel, L. and Lin, C.T., 2005. Expectations of patients and physicians regarding patient-accessible medical records. *JMIR*, 7(2), p.e13.
- [5] Wildemuth, B.M., Blake, C.L., Spurgin, K., Oh, S. and Zhang, Y., 2006. Patients' perspectives on personal health records: An assessment of needs and concerns. *Critical Issues in eHealth Research*, 2006.
- [6] Hassol, A., Walker, J.M., Kidder, D., Rokita, K., Young, D., Pierdon, S., Deitz, D., Kuck, S. and Ortiz, E., 2004. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. *Journal of the American Medical Informatics Association*, 11(6), pp.505-513.
- [7] Arora, N.K., Johnson, P., Gustafson, D.H., McTavish, F., Hawkins, R.P. and Pingree, S., 2002. Barriers to information access, perceived health competence, and psychosocial health outcomes: test of a mediation model in a breast cancer sample. *Patient education and counseling*, 47(1), pp.37-46.
- [8] Pratt, W., Unruh, K., Civan, A. and Skeels, M.M., 2006. Personal health information management. *Communications of the ACM*, 49(1), pp.51-55.
- [9] Samba, A.V., Mansour, E., Hawke, S., Zereba, M., Greco, N., Ghanem, A., Zagidulin, D., Aboulmaga, A. and Berners-Lee, T., Solid: A Platform for Decentralized Social Applications Based on Linked Data.
- [10] Shaban-Nejad, A., Lavigne, M., Okhmatovskaia, A. and Buckeridge, D.L., 2017. PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. *Annals of the New York Academy of Sciences*, 1387(1), pp.44-53.
- [11] Brenas, J.H., Shin, E.K. and Shaban-Nejad, A., Adverse Childhood Experiences Ontology for Mental Health Surveillance, Research, and Evaluation: Advanced Knowledge Representation and Semantic Web Techniques. *JMIR Ment Health*. 2019 May 21;6(5):e13498.
- [12] Al Manir M.S., Brenas, J.H., Baker, C.J., Shaban-Nejad, A. A Surveillance Infrastructure for Malaria Analytics: Provisioning Data Access and Preservation of Interoperability. *JMIR Public Health Surveill*. 2018 Jun 15;4(2):e10218.

The SmartSight Project: Use of Electronic Glasses to Improve Impaired Fields of Vision

Berglind F. SMARADOTTIR^{a,1}, Niels F. GARMANN-JOHNSEN^{b,c}, Steinar OMNES^d
Ann-Elisabeth LUDVIGSEN^e and Harald REISO^{d,f}

^a *Department of Information and Communication Technology, University of Agder, Norway*

^b*Department of Society, NORCE Norwegian Research Centre, Norway* ^c*Department of Information Systems, University of Agder, Norway* ^d*Research Department, Sørlandet Hospital Trust, Norway* ^e*Department of eHealth and public sector, Egde Consulting, Norway*

^f*General Practice Research Unit (AFE), Department of General Practice, Institute of Health and Society, University of Oslo, Norway*

Abstract. Electronic glasses use advanced assistive technology that can improve function for persons with visual impairments. This paper presents work in progress in the SmartSight project, where existing versions of electronic glasses are adapted and tested for persons with the visual impairments; macular degeneration and retinitis pigmentosa. The project aims are to adapt and innovate electronic glasses for persons with impaired fields of vision and study the impact on the daily function and quality of life. In a pilot test of the electronic glasses, promising results were found that imply a big difference for persons with impaired vision. The outcome of the SmartSight project might change the clinical practise on treatment and rehabilitation of persons with impaired fields of vision, and on how the support services are organised.

Keywords. E-accessibility, Visual impairment, Augmented reality, E-glasses, Usability evaluation

1. Introduction

Two large patient groups have eye diseases that impair their fields of vision, inducing problems in everyday living and thus reduced function. Macular degeneration (MD) causes the central vision to deteriorate while the peripheral vision remains intact [1], see Figure 1 middle. One in ten older than 70 years of age have functional impairments due to MD [2]. In retinitis pigmentosa (RD) central and peripheral fields of vision are impaired, giving a “tunnel vision” for persons affected [3], see Figure 1 right. This is problematic for orientation and implies more impaired vision in dark surroundings [4]. The total prevalence of all different forms of retinitis pigmentosa is variably reported (one case for each 2500 - 7000 persons), but represents one of the most common causes of blindness or severe low-vision in people 20 to 60 years old [5].

¹ Corresponding author, e-mail: berglind.smaradottir@uia.no



Figure 1. The field of vision for normal vision (left), macular degeneration (middle) and retinitis pigmentosa (right).

In this context, the SmartSight project (*Electronic glasses to improve function for persons with impaired fields of vision*) addresses research gaps connected to clinical studies of eHealth assistive technology aiming to improve function and quality of life for persons with impaired fields of vision. The software in existing technology of electronic glasses (eGlasses) will be adapted to improve function and coping for persons with impaired fields of vision. The effects of using eGlasses will be measured and tested for persons with macular degeneration and retinitis pigmentosa.

The main hypotheses are that use of electronic glasses will have beneficial effects on individual, organisational and societal levels. The research questions (RQs) stated for the SmartSight project:

RQ1: How can the functionality of existing eGlasses products be adapted to improve function and coping for persons with impaired fields of vision?

RQ2: What benefits and effects of using such customised products can be measured for individuals and for the society?

RQ3: What are the user experiences with such customised products?

2. Methodology

To address the user experiences, a mixed-methods research approach will be applied to evaluate the usefulness and user satisfaction with the eGlasses [6][7]. Initially, pilot tests will be made. User evaluations led by a cross-disciplinary research team, will be performed in a usability laboratory together with persons affected by macular degeneration and retinitis pigmentosa, followed by user satisfaction questionnaires and qualitative interviews [8]. By this systematic approach, real end-users get the opportunity to share their user experiences regarding the eGlasses and the results will be used for iterative refinements of the eGlasses. For each iteration in the technical development, 4- 6 test participants evaluate the functionality and usability of the eGlasses. All laboratory tests will be video-recorded for analysis purposes [9][10].

To evaluate the long-term experiences; a 6-12 months long field study will be made with 10-12 participants using a high-fidelity prototype of eGlasses. During the study, there will be regular contacts between the research team and the test participants for evaluating impacts on fields of vision and daily life. Recruitment of test participants is made in close collaboration with The Norwegian Association of the Blind and Partially Sighted [11], the Eye Clinic at Sørlandet Hospital Trust [12] and The National Centre for Optics, Vision and Eye Care, University of South-Eastern Norway [13].

3. Results

The SmartSight project is in an early study phase. A workshop was organised together with The Norwegian Association of the Blind and Partially Sighted, where pilot tests of eGlasses were made. The eGlasses had mounted cameras, and modified pictures from the cameras were projected in the glasses.



Figure 2. The pilot test of different types of eGlasses. Photo: Mona Hauglid.

The tests showed promising results using unmodified eGlasses that may imply great differences for persons with retinitis pigmentosa [14][15], see Figure 3. Unmodified eGlasses did not give any benefit for persons with macular degeneration, but the modified versions are promising, see Figure 4.

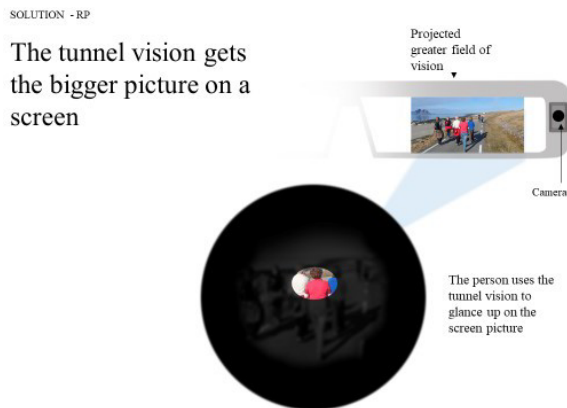


Figure 3. A model of principle function of the eGlasses for retinitis pigmentosa.

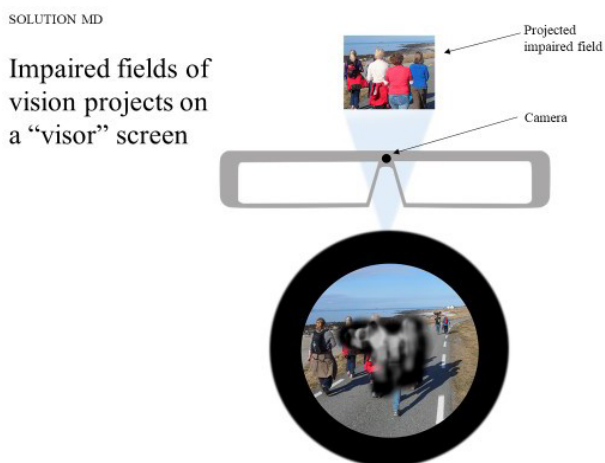


Figure 4. A model of principle function of the eGlasses for macular degeneration.

4. Conclusion

The main contributions of the SmartSight study lie in the adaption and innovation of existing eGlasses technology, a detailed analysis of their usefulness, and user experiences from persons with visual impairments caused by macular degeneration and retinitis pigmentosa. These outcomes might change the clinical practise on examination and rehabilitation for persons with impaired field of vision, and improve quality of life by increasing employability and participation in social life. The outcomes might also have impact on the organisation of support services connected to this area of concern.

4.1 Acknowledgement

Thanks to The Norwegian Association of the Blind and Partially Sighted and to the persons that voluntarily participated in the pilot test. This study was funded by Research Grant no. 16/00528-27 of December 21st, 2016, from the South-Eastern Regional Health Authority in Norway.

References

- [1] A.C. Bird, N.M. Bressler, S.B. Bressler, I.H., Chisholm, G. Coscas, M.D. Davis, et.al., An international classification and grading system for age-related maculopathy and age-related macular degeneration, *Surv Ophthalmol* **39**:5 (1995) 367-374, doi:[https://doi.org/10.1016/S0039-6257\(05\)80092-X](https://doi.org/10.1016/S0039-6257(05)80092-X)
- [2] J.M. Colijn, G.H. Buitendijk, E. Prokofyeva, D. Alves, D. M.L. Cachulo, A.P. Khawaja et.al., Prevalence of age-related macular degeneration in Europe: the past and the future, *Ophthalmology* **124**:12 (2017), 1753-1763, doi: 10.1016/j.ophtha.2017.05.035
- [4] C. Hamel, Retinitis pigmentosa, *Orphanet J. Rare Dis* **1**:1, (2006), 40, doi:10.1186/1750-1172-1-40

- [5] D.T. Hartong, E.L. Berson, T.P. Dryja, Retinitis pigmentosa, *The Lancet* **368**:9549 (2006), 1795-1809, doi:[https://doi.org/10.1016/S0140-6736\(06\)69740-7](https://doi.org/10.1016/S0140-6736(06)69740-7)
- [6] F. Parmeggiani, Clinics, epidemiology and genetics of retinitis pigmentosam, *Curr Genomics* **12**:4 (2011), 236–237. doi:10.2174/138920211795860080
- [7] J.W. Creswell, V.L.P. Clark, *Designing and conducting mixed methods research*. SAGE Publications Inc, 2007.
- [8] R.B. Johnson, A.J. Onwuegbuzie, L.A. Turner, Toward a definition of mixed methods research, *J Mix Methods Res* **1**:2 (2007), 112-133.
- [9] M.W. Jaspers, A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence, *Int J Med Inform* **78**:5 (2009), 340-353.
- [10] B. Smaradottir, J. Håland, S. Martinez, Å.R Somdal, R.W. Fensli, Recommendations on a test infrastructure for evaluation of touchscreen assistive technology for visually impaired users, Linköping University Press, 2015.
- [11] B.F. Smaradottir, J.A. Håland, S.G. Martinez, User Evaluation of the Smartphone Screen Reader VoiceOver with Visually Disabled Participants, *Mob Inf Syst* (2018), Article ID 6941631, doi: <https://doi.org/10.1155/2018/6941631>
- [12] The Norwegian Association of the Blind and Partially Sighted. [cited 2020 Jan 1]. Available from: <https://www.blindeforbundet.no/om-blindeforbundet/brosjyrer/muligheter-til-et-aktivt-liv-engelsk>
- [13] The Eye Clinic at Sørlandet Hospital Trust. [cited 2020 Jan 1]. Available from: <https://sshf.no/avdelinger/somatikk-arendal/oye/oyepoliklinikk>
- [14] National Centre for Optics, Vision and Eye Care, University of South-Eastern Norway. [cited 2020 Jan 1]. Available from: <https://www.usn.no/english/research/our-research/vision-science/national-centre-for-optics-vision-and-eye-care/>
- [15] B. Borgersen J. Claesson, M. Grundetjern, *Adapting smart glasses for persons with loss of peripheral vision*. Bachelor thesis, University of Agder, 2018. [cited 2020 Jan 1]. Available from: https://flåttsenderet.no/wp-content/uploads/2019/04/e-Glasses_bachelor_report_UiA_140518.pdf
- [16] K. Lorentzen, *Manipulation of virtual pictures to compensate impaired fields of vision (In Norwegian Manipulering av virtuelle bilder for kompensering av tapt synsfelt)*, Bachelor thesis, University of Agder, 2017. [cited 2020 Jan 1]. Available from: https://flåttsenderet.no/wp-content/uploads/2019/04/e-Glasses_Kenneth_Lorentzen_Short-form_Bachelorreport_UiA_sept_2017.pdf

Three-Dimensional Volumetric Renal Reconstruction Based on Geometrical Coefficients

Tomasz LES^{a,1}, Tomasz MARKIEWICZ^{a,b}, Mirosław DZIEKIEWICZ^b and
Malgorzata LORENT^b

^aWarsaw University of Technology, Warsaw, Poland

^bMilitary Institute of Medicine, Warsaw, Poland

Abstract. The article presents an innovative method of 3D computer tomography (CT) image reconstruction of kidney. Diagnosis based on CT scanning allows to obtain projections of multi-dimensional object, made from different directions in order to create cross-sectional (2D) slices. Standard techniques for identifying kidneys in CT images analyze each 2D slice separately. It causes different reconstruction accuracy for the same object at its different heights. This is the main problem of a machine-learning systems. Reconstruction error of end-slices of the kidney model is often greater than the error of the kidney's middle part. The main idea of the technique presented in this paper is to analyze the largest coherent 3D spatial-areas. This technique allows to increase the accuracy of kidney detection as well as to decrease the FP (false positive) error. An additional advantage of the developed algorithm is the possibility of obtaining a precise model representing the 3D view of an entire kidney.

Keywords. renal, reconstruction, geometrical-coefficients

1. Introduction

Quick and effective diagnosis of kidney diseases is possible by usage of a precise X-ray imaging. Increasingly more accurate projection, combined with novel methods of an automatic segmentation and visualization of the organs, e.g. kidneys, allows a better diagnosis of the disease and also a selection of a more appropriate treatment. An accurate kidney detection is particularly important for the cancer diagnosis. According to WHO Cancer Mortality Database, 400 000 new cases were diagnosed in 2018 [1]. Kidney cancer is the ninth most common cancer among men and 14th among women. The highest incidence rate covers the countries of Central and Eastern Europe (Belarus, the Czech Republic) and the Baltic States (Lithuania, Latvia and Estonia).

Current techniques for diagnosis of kidney diseases are based on CT image analysis, followed by manual or semi-automatic segmentation in order to detect the kidney area. Although, a 3D reconstruction of the kidney is often used in diagnosis for a better visualization. Modern CT scanners generate large number of scans with high density. This is an advantage, however, it causes also that the manual analysis,

¹ Corresponding Author, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland;
E-mail: lest@ee.pw.edu.pl.

performed by specialists, of a large number of images, becomes time consuming and involves the risk of human error caused by, for example, tiredness. Therefore, there is a necessity to develop a new, more accurate and faster techniques for an automatic diagnosis and visualization of kidneys. Current kidney identification techniques are based on the analysis of each slice separately. The consequence is that the accuracy of recognition of an individual kidney contours on different slices can vary significantly. This is particularly important when analyzing the end-slices of a kidney where the accuracy of a detection is often worse than the accuracy of a kidney detection in the middle slices. This causes false positive error, difficult to decrease with standard techniques.

In this article, a novel volumetric technique of a renal reconstruction is presented, based on an analysis of geometrical coefficients for 3D kidney model. Renal contour detection is performed on all scans together unlike in many popular methods which analyses each slice separately. The final 3D model generated using our technique is also more precise than model using standard techniques.

2. Automatic renal segmentation technique

In this section we present a complete system of the 3D kidney reconstruction. The starting point of the algorithm is a set of n images encoded as gray-scale images, representing individual slices of renal. Input images can be obtained, e.g. using morphological operations [2,3], segmentation methods [4-6] or methods based on machine learning [7-9]. Regardless of the used method, the key task is to calculate the appropriate threshold for binarization final mask. In our experiment we used U-Net multilayer network model to obtain the input data: raw CTA images of dimensions of 512 x 512, frame size 48x48 (defined the input layer of U-Net), encoder depth equals 4 and convolution filter size equals 3. Other neural network parameters we applied: initial learning rate: 0.07, momentum 0.9 and LZ regularization 0.0002, number of epochs: 20 and minimal batch size: 256.

Output of the network frame classification is the gray-scale image. The value of each pixel $c = [x, y]$ equals the number of votes that the pixel c belongs to the renal area when the sequential frames are classified by the network.

It is difficult to calculate the optimal threshold value (t value) for a segmentation of renal area. Moreover, t value differs for different cases and also for each renal in one case. The main, novel contributions of this work include a method of calculating geometric coefficients for the entire 3D renal model, for different threshold levels. Following coefficients are calculated for each renal separately: fc - the circularity ratio (1) and fh - the compactness factor (2).

$$fc = \frac{A}{P} \quad (1),$$

$$fh = \frac{P^2}{A} \quad (2)$$

Directly measured parameters are denoted by the following symbols:

- A – real area of the renal
- P – perimeter of the renal

Finally, the p factor is calculated by the formula (3):

$$p = \frac{fc}{fh}$$

(3)

Next, the algorithm selected minimal p value and corresponding t value. Table 1 presents the dependence of the p value and t value, as well as of the fc and fh factors for one case.

Table 1. Dependence of the t , p , fc and fh factors for one case.

t	fc	fh	p	t	fc	fh	p
17	10147.06	7.92	1281.095	28	2063.42	12.03	171.474
20	9868.72	7.80	1265.137	29	2071.86	11.98	172.970
22	9756.89	7.71	1265.934	30	2102.20	11.86	177.223
23	2177.78	11.91	182.855	31	2116.63	11.79	179.543
24	1987.00	12.39	160.358	40	2302.47	11.02	209.019
25	1983.68	12.37	160.356	60	4774.24	5.88	812.425

Fig. 1 presents plot of minimal p value range in relation to t value.

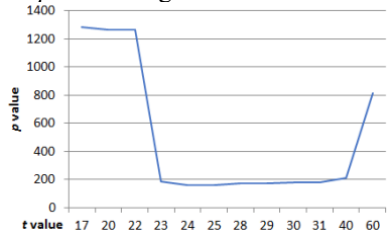


Figure 1. Plot of minimal p value range in relation to t value.

The minimal p value is obtained for t range from 23 to 40. For these values the reconstruction is the most precise. Our method allows a simple selecting the minimum value of p (which is equal: 160.356 in this case) and an optimal threshold value t (equals 25 in this case). The more general analysis of other cases allows to observe that the optimal t value ranges from 20 to 50. This technique is applied for each renal separately. Fig. 2 presents the results of a complete reins reconstruction using t value equals: [22,25,60].

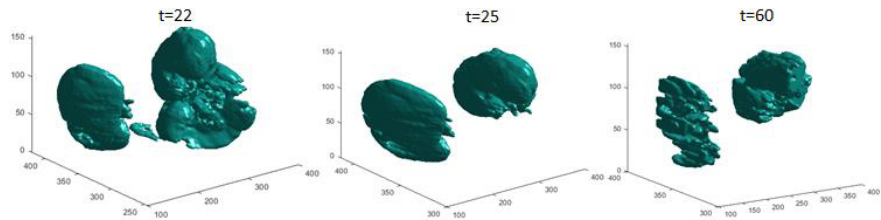


Figure 2. Complete reconstruction of reins for t value of: [22,25,60], for one patient.

In this case, the best reconstruction quality can be obtained for the threshold $t = 25$. This value was calculated automatically basing on coefficient factors fc , fh and p value.

A full diagram of a complete system for 3D reconstruction of the reins is shown in fig 3.

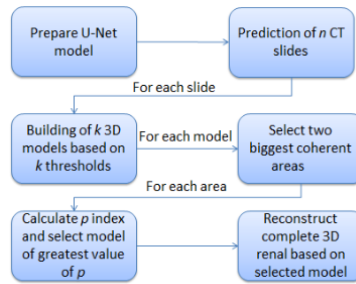


Figure 3. A full diagram of a complete system for 3D reconstruction of reins.

3. Test and results

In numerical experiments we have recognized 21 cases, each consisting of 60-250 slices. In total, 1335 CT images were tested. For each case, automatically, a complete 3D reconstruction of a renal was realized as well as the final binary masks for each individual slice were calculated. Each mask was compared to mask obtained by an expert from the Military Institute of Medicine and final Renal Sørensen-Dice coefficient was calculated. Detailed numerical results of recognition accuracy are presented in Table 2.

Table 2. The accuracy of renal area recognition for 21 cases.

Case	Accuracy	Case	Accuracy	Case	Accuracy
1	87.60	8	84.82	15	80.41
2	92.45	9	59.91	16	85.66
3	75.97	10	90.08	17	87.02
4	90.70	11	78.36	18	91.02
5	88.85	12	89.48	19	94.41
6	96.60	13	94.29	20	95.24
7	91.00	14	89.39	21	95.13
mean					87.54

Experimental results show the good performance of the proposed system. The main accuracy of kidney reconstruction equals 87.54%.

The results of renal boundary detection for exemplary slices were presented on fig 4. The green line represents the expert's result and the red line represents the system's result.

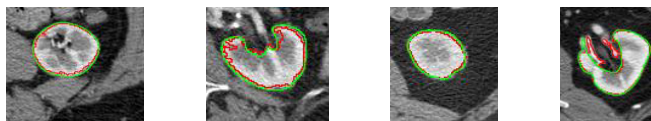


Figure 4. Results of renal boundary detection for exemplary slices. The green line represents the expert's outlines and the red line represents the system's results.

4. Conclusion

This paper presents a novel method of the 3D renal reconstruction. The presented technique allows to obtain a precise boundary of the reins in both three-dimensional

and two-dimensional projections. Our algorithm allows for calculating the optimal thresholding value for gray-scale images. An automatic selection of threshold is a difficult task due to the very high variability of the intensity of each slice. The existing techniques for visualizing the correlation of threshold values and reconstruction results use the roc curve. However, those techniques do not allow for automatic selecting of one optimal value. The proposed technique was tested using a set of data generated by the U-Net network, however, many other techniques generating gray-scale images (e.g. basing on morphological operations or segmentation methods), can be used in our solution. The main novel idea used in the presented system is to analyze geometric coefficients (circularity and compactness) as volumetric factors. The result of the coefficient calculation allows selecting optimal threshold value, binarization and final renal boundary recognition. In order to assess the quality of the proposed system, numerical tests were performed. Each mask was compared to a mask obtained by an expert from a medical institute. The numerical results of this verification have been presented and discussed in this paper. The measurable benefit of the presented system is a practical tool for visualization and evaluation of very significant cancerous change. Despite satisfactory research results, it is necessary to constantly develop new methods to support medical diagnostics in order to improve treatment results.

Acknowledgment

This work has been supported by the National Science Centre (2016/23/B/ST6/00621 grant), Poland

References

- [1] <http://www-dep.iarc.fr/WHODb/WHODb.htm>
- [2] Les, T., Markiewicz, T., Dziekiewicz, M. Lorent, M.; Automatic recognition of the kidney in CT images ,19th International Conference Computational Problems of Electrical Engineering, 2018, IEEE
- [3] Saii, M., Kraitem, Z., Automatic Detection and Segmentation of Kidneys in Magnetic Resonance Images Using Image Processing Techniques. Biomedical Statistics and Informatics. Vol. 2, No. 1, 2017, pp. 22-26. doi: 10.11648/j.bsi.20170201.15
- [4] Song, H., Kang, W., Zhang, Q., Wang, S., Kidney segmentation in CT sequences using SKFCM and improved GrowCut algorithm, IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2014), Belfast, UK. 2-5 November 2014
- [5] Daw-Tung, L., Chung-Chih, L., Siu-Wan, H., Computer-Aided Kidney Segmentation on Abdominal CT, IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 10, NO. 1, January 2006
- [6] Yang, G., Gu, J., Chen, Y., Liu, W., Tang, L., Shu, H., Toumoulin, C.: Automatic kidney segmentation in ct images based on multi-atlas image registration. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE (2014)
- [7] Cicek, O., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp.424-432. Springer (2016)
- [8] Cuingnet R., Prevost R., Lesage D., Cohen L.D., Mory B., Ardon R. Automatic Detection and Segmentation of Kidneys in 3D CT Images Using Random Forests. In: Ayache N., Delingette H., Golland P., Mori K. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. MICCAI 2012. Lecture Notes in Computer Science, vol 7512. Springer, Berlin, Heidelberg
- [9] Kutikov, A., Uzzo, R.G. The renal nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. The Journal of urology 182(3), (2009)

Towards Precise Descriptions of Medical Free/Libre and Open Source Software

Franziska JAHN^{a,1}, Michelle BINDEL^a, Konrad HÖFFNER^a, Maryam GHALANDARI^a, Birgit SCHNEIDER^a, Sebastian STÄUBERT^a, Verena DORNAUER^b, Thomas KAROPKA^c, Elske AMMENWERTH^b, Alfred WINTER^a
^a*Institute of Medical Informatics, Statistics and Epidemiology, Leipzig University*
^b*Institute of Medical Informatics, UMIT – Private University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria,*
^c*Medfloss.org, Germany*

Abstract. The web portal Medfloss.org lists over 360 medical free/libre and open source software (MEDFLOSS) projects. These projects are described with the help of a self-developed nomenclature. Due to inconsistencies, the nomenclature shall be replaced by HITO, the Health IT Ontology. HITO is developed iteratively based on different use cases. This paper aims to describe methods and results of the second HITO use case in which HITO is extended to improve the description, retrieval and comparisons of MEDFLOSS projects on Medfloss.org. We use a mixed-methods approach to add concepts and relationships to describe MEDFLOSS precisely. The resulting HITO version stresses functional descriptions based on features and supported enterprise functions, rather than just describing technical characteristics. However, describing a larger number of MEDFLOSS projects requires the commitment of the community.

Keywords. Health Information Systems, Software, Semantics

1. Introduction

Medical free/libre and open source software (MEDFLOSS) is widely used in academia and in low-resource settings, but also in healthcare institutions of industrial countries such as the US and UK [1]. The web portal Medfloss.org gives an overview of 363 MEDFLOSS projects, 111 professional service providers and 436 publications dealing with MEDFLOSS [2]. The portal is provided in cooperation with the FLOSS working groups of the European Federation of Medical Informatics, the International Medical Informatics Association and the International Society for Telemedicine & eHealth. About 3.350 visitors consult the website every month.

To describe software projects, Medfloss.org uses a self-developed nomenclature describing software by categories such as “application type”, supported “enterprise functions” and “license”. Since the establishment of Medfloss.org in 2010, the uncontrolled growth of the nomenclature has led to inconsistencies such as misassignments of descriptors to characteristics (e.g. “Virtual Reality” is listed as

¹ Franziska Jahn, Leipzig University, Medical Faculty, Institute of Medical Informatics, Statistics and Epidemiology, Härtelstr. 16-18, 04107 Leipzig, E-mail: franziska.jahn@imise.uni-leipzig.de

enterprise function) and the use of synonyms and homonyms for the descriptions of MEDFLOSS projects.

The Austrian-German research project HITO aims at the development of a **Health IT Ontology** for the systematic description of application systems and software products used in healthcare [3]. Whereas software products are tailor-made or purchased pieces of software that can be installed on a computer system, application systems are already installed software products that are adapted to a certain organizational environment [4].

HITO is developed iteratively by means of five use cases that currently lack precise application system and software product descriptions. Each use case is intended to add new concepts and relationships to HITO. In the first use case, we developed HITO V1 in order to improve the retrieval of evaluation studies examining application systems in healthcare settings [5]. HITO V1 systematically describes evaluated application systems by features, user groups and organizational units.

In the second use case, HITO V1 is extended by concepts and relationships for the precise description of MEDFLOSS (HITO V2) to improve retrieval and comparisons of MEDFLOSS projects on Medfloss.org.

This paper aims to describe methods and results of developing HITO V2.

2. Methods

We used a mixed-methods approach to identify important categories for the description of MEDFLOSS. The methods described in sections 2.1 to 2.4 were applied in parallel.

2.1. Surveying the Medfloss.org portal operators

In semi-structured interviews, two of the Medfloss.org web portal operators were asked about their experiences with the current terminology used on Medfloss.org and their ideas for improvement. On a 4-point scale ranging from “unimportant” to “very important”, both rated 8 out of 11 categories of the current terminology (enterprise function, application type, status, license, standard, language, client type, and platform) as “important” or “very important” for the description of the projects. The categories “popularity”, “database” and “programming language/toolkit” were each rated “less important” and “important” by one of the interviewees. They pointed out that the importance of these categories also depends on the perspective of the MEDFLOSS portal user. While health professionals are interested in functional software descriptions, software developers are more interested in technical categories.

However, from their point of view, the terminology currently lacks possibilities to describe the functionalities and application types more precisely. Therefore, the integration of existing terminologies such as WHO’s classification of digital health interventions [6] should be considered.

2.2. Comparison of three terminologies for describing MEDFLOSS functionalities

During the first HITO use case, we identified existing health IT taxonomies and ontologies [5]. To describe MEDFLOSS functionalities, we selected three of these taxonomies, which cover a broad range of health enterprise functions or application

system types systems, for a deeper analysis: (1) The reference model for the domain layer of hospital information systems (RDHIS) provides a hierarchy of enterprise functions (e.g. “medical admission”, “execution of nursing procedures”) [7]. Although it is originally developed for hospitals, many of these enterprise functions also fit in other healthcare settings. (2) The textbook “Health Information“ includes a list of application system types (e.g. radiology information system, patient administration system) used in healthcare and tries to avoid conceptual overlappings [4]. (3) Varshney et al. systematically developed a taxonomy of healthcare applications consisting of eight application system types [8].

For 356 projects listed on Medfloss.org, we counted how many of them could be described by the three taxonomies’ classes. The RDHIS gained best results in this comparison (Table 1). 71 % of the MEDFLOSS projects could be described by it. Only 13 of the 68 RDHIS’ enterprise functions could not be used for any description. With the help of application system list and the taxonomy of healthcare applications, only 42 % and 29 % of the projects could be described, respectively.

Table 1. Comparison of three taxonomies to describe 356 projects listed on Medfloss.org

	(1) Reference model for the domain layer of HIS [7]	(2) Application system list [4]	(3) Taxonomy of healthcare applications [8]
Number of classes	68	16	8
Number of Medfloss projects assignable to classes (n=356)	253 (71%)	151 (42%)	105 (29%)

2.3. Analyzing the Architecture of a MEDFLOSS-based Laboratory Environment

The MI-Lab at Leipzig University is a small computerized hospital information system architecture built up from several MEDFLOSS components [9]. Its analysis reveals that MEDFLOSS products can be provided as parts of other MEDFLOSS products. The Bahmni software used as electronic medical record (EMR) system in MI-Lab is composed of four other MEDFLOSS products providing a laboratory information system (LIS), an enterprise resource planning system (ERP), a patient management system and a picture archiving and communication system (PACS).

2.4. MEDFLOSS Literature Review

In order to find important characteristics for the description of MEDFLOSS products, we conducted a snowball search starting from a review on MEDFLOSS publications [1] which was complemented by a Pubmed and Google Scholar search on open source software characteristics and comparisons. Four out of seven articles providing structured descriptions of MEDFLOSS [10–13] list characteristics that are worth to be considered as HITO classes. However, the use of different self-developed lists of features or functionalities in the articles demonstrates the lack of a consented or at least organized terminology, but also unclear borders between features (e.g. “templates”) and supported enterprise functions (e.g. “billing”). Most of the categories used by Medfloss.org (2.1) can also be found in the articles: license, database, platform, programming language, language. Further categories describing privacy, security, usability, performance, customizability, user and developer support issues should be considered for inclusion into HITO.

3. Results

Based on our mixed-methods approach, we made the following design decisions for HITO V2 and explain them referring to the well-documented Bahmni software [14]:

- (A) All categories for the description of MEDFLOSS used on Medfloss.org are integrated into HITO V2 as classes because of their importance rating (2.1) and their use in MEDFLOSS literature (2.4). For example, Bahmni uses the license “AGPL 3.0” and supports the platform “CentOS”.
- (B) Functionalities of software products are described by combinations of supported enterprise functions and features (2.2, 2.4), e.g., Bahmni supports the enterprise function “Patient Registration” for which it provides the features “Register patients with their demographics”, “Identification contact”, etc. In a further step, these citations of software descriptions can be linked to entries of established catalogues for enterprise functions (e.g. [6,7]), and features (e.g. [6]), respectively. The enterprise function “Patient Registration” can be linked to the RDHIS function “Administrative Patient Admission”, for example. See Figure 1 for the corresponding HITO classes and relationships.

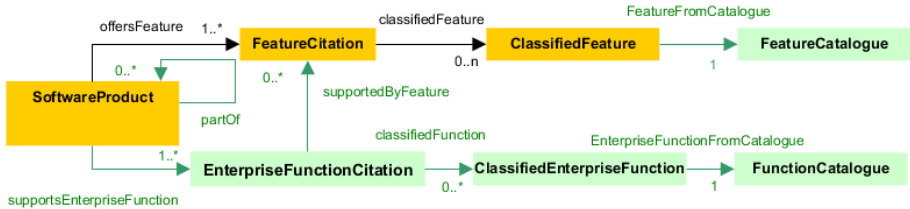


Figure 1. HITO V2: Relationships between functions and features as UML class diagram. Every citation of a MEDFLOSS description can be linked to a “classified” entry of a catalogue.

- (C) The RDHIS is transformed into a reference model for the domain layer of health information systems which includes enterprise functions for specific diagnostics and therapies, mobile health, research and public health (2.2).
- (D) The class “Software product” is linked with itself by a “part of” relationship (2.3). Thus, the relationship “Open MRS is part of Bahmni” can be expressed. HITO V2 is available as Linked Open Data as RDF on <https://hitontology.eu> [3].

4. Discussion and Conclusion

Using a mixed-method approach, HITO could be extended by a variety of classes for the description of MEDFLOSS projects. Besides rather technical characteristics like platform or programming language, the precise description of software functionalities for medicine and health care plays an important role for the selection and comparison of MEDFLOSS. Thus, the domain-specific description abilities of HITO can be considered as one of its strengths and could, in the future, make it comparable to the EDAM ontology used in bioinformatics to describe operations and data [15]. However, due to the different terminologies for functionalities and application system types already in use, a unification of terms cannot be achieved in the short term. The HITO

project therefore integrates different existing terminologies for features and enterprise functions with the long-term goal of linking their terms.

Although a user searching for a specific software often relies on allegedly established terms for application system types (e.g. EHR, EMR, PACS, LIS), these terms are not clearly defined. For example, an “electronic health record (EHR) software” might be developed for the use in practices, hospitals or healthcare networks. It may be the electronic collection of patients’ documents or it may provide full medical process support. For this reason, HITO currently focuses on functional descriptions. After describing a large number of software products, clustering might help to improve definitions of application system types. However, coming to a huge amount of meaningful MEDFLOSS descriptions by HITO depends on good documentation of the software projects and the commitment of the community.

Acknowledgements

The HITO project is funded by the Austrian Science Fund FWF (I3726-N31) and the German DFG fund (WI 1605/11-1).

References

- [1] T. Karopka, H. Schmuhl, and H. Demski, Free/Libre open source software in health care: a review. *Healthc Inform Res* 20 (2014), 11–22.
- [2] H. Streidl, H. Demski, and T. Karopka, Medical Free/Libre and Open Source Software, <https://www.medfloss.org> [cited 2019 September 25].
- [3] HITO Team, HITO: A Health IT Ontology for systematically describing application systems and software products in health IT, <https://hitontology.eu> [cited 2019 September 25].
- [4] A. Winter, R. Haux, E. Ammenwerth, B. Brigl, N. Hellrung, and F. Jahn, *Health Information Systems: Architectures and Strategies*, Springer, London, 2011.
- [5] E. Ammenwerth, V. Dornauer, M. Ghalandari, F. Jahn, N. de Keizer, and A. Winter, An Ontology for Describing Health IT Interventions: Methodological Considerations. *Stud Health Technol Inform* 264 (2019), 1419–1420.
- [6] WHO, *Classification of digital health interventions: A shared language to describe the uses of digital technology for health*, Geneva, 2018.
- [7] G. Hübner-Bloder, E. Ammenwerth, B. Brigl, and A. Winter, Specification of a Reference Model for the Domain Layer of a Hospital Information System. In *Proceedings of MIE2005*, R. Engelbrecht, ed. IOS Press, Amsterdam, Washington, DC, 2005, pp. 497–502.
- [8] Varshney, U., Nickerson, RC and J. Muntermann, Taxonomy Development in Health-IT. In *AMCIS 2013 Proceedings*, 2013.
- [9] K. Brandt, M. Löbe, M. Schaaf, F. Jahn, A. Winter, and S. Stäubert, MI-Lab - A Laboratory Environment for Medical Informatics Students. *Stud Health Technol Inform* 228 (2016), 48–52.
- [10] A.E. Flores Zuniga, K.T. Win, and W. Susilo, Functionalities of free and open electronic health record systems. *Int J Technol Assess Health Care* 26 (2010), 382–389.
- [11] A.A. Zaidan, B.B. Zaidan, A. Al-Haiqi, M.L.M. Kiah, M. Hussain, and M. Abdalnabi, Evaluation and selection of open-source EMR software packages based on integrated AHP and TOPSIS. *Journal of Biomedical Informatics* 53 (2015), 390–404.
- [12] S. Purkayastha, R. Allam, P. Maity, and J.W. Gichoya, Comparison of Open-Source Electronic Health Record Systems Based on Functional and User Performance Criteria. *Healthc Inform Res* 25 (2019), 89–98.
- [13] P.S. Millard, J. Bru, and C.A. Berger, Open-source point-of-care electronic medical records for use in resource-limited settings: systematic review and questionnaire surveys. *BMJ Open* 2 (2012).
- [14] Bahmni Coalition, Hospital System for Low Resource Settings, <https://www.bahmni.org> [cited 2019 September 27].

- [15] J. Ison, M. Kalas, I. Jonassen, D. Bolser, M. Uludag, H. McWilliam, J. Malone, R. Lopez, S. Pettifer, and P. Rice, EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics* 29 (2013), 1325–1332.

Trajectories of Disease Accumulation Using Electronic Health Records

Pere PLANELL-MORELL ^{a,1}, Madhavi BAJEKAL ^b, Spiros DENAXAS ^c,
Rosalind RAINE ^b, and Daniel C. ALEXANDER ^a

^aCentre for Medical Image Computing, University College London

^bDepartment of Applied Health Research, University College London

^cInstitute of Health Informatics, University College London

Abstract. Multimorbidity is a major problem for patients and health services. However, we still do not know much about the common trajectories of disease accumulation that patients follow. We apply a data-driven method to an electronic health record dataset (CPRD) to analyse and condense the main trajectories to multimorbidity into simple networks. This analysis has never been done specifically for multimorbidity trajectories and using primary care based electronic health records. We start the analysis by evaluating temporal correlations between diseases to determine which pairs of disease appear significantly in sequence. Then, we use patient trajectories together with the temporal correlations to build networks of disease accumulation. These networks are able to represent the main pathways that patients follow to acquire multiple chronic conditions. The first network that we find contains the common diseases that multimorbid patients suffer from and shows how diseases like diabetes, COPD, cancer and osteoporosis are crucial in the disease trajectories. The results we present can help better characterize multimorbid patients and highlight common combinations helping to focus treatment to prevent or delay multimorbidity progression.

Keywords. multimorbidity, electronic health records, disease trajectory, multiple disease progression model

1. Introduction

Multimorbidity (the coexistence of two or more chronic diseases in an individual) is a growing health and healthcare challenge. There is significant evidence showing that the prevalence of patients with multiple chronic conditions has increased over time [1]. Nevertheless, we still do not know the temporal dynamics of disease accumulation over time in the development of multimorbidity. In this paper we study which are the trajectories of disease accumulation that patients follow to become multimorbid. Understanding what are the characteristics of these trajectories can be helpful to better identify and predict the disease progression of patients. We adapt the methodology proposed by Jensen et al [2] and we apply it to primary care electronic health records (EHRs) instead of secondary care EHRs. If we want to study multimorbidity we need to use primary care data since most chronic condition are treated in general practice. However, it is challenging

¹Corresponding Author: Pere Planell-Morell; E-mail: p.morell@ucl.ac.uk

Table 1. List of chronic conditions in scope

Atrial Fibrillation (AF)	Anxiety	Alcohol problems
COPD	Coronary Heart Disease (CHD)	Chronic Kidney Disease (CKD)
Cancer	Chronic liver disease (CLD)	Depression
Diabetes	Dementia	Diverticulitis of intestine (Div)
Epilepsy	Glaucoma	Hypothyroidism
Heart Failure (HF)	Inflammatory Bowel Disease (IBD)	Learning Disabilities (LD)
Multiple Sclerosis (MS)	Motor Neurone Disease (MND)	Osteoporosis (Osteo)
Osteoarthritis	Peripheral Arterial Disease (PAD)	Parkinson’s disease
Psychoactive substance misuse (Substance)	Prostate disorders	Psoriasis
Rheumatoid Arthritis (RA)	Stroke or TIA	Severe Mental Illness (SMI)

to work with primary care data because it is not standardized between countries and the disease coding system used in the UK cannot be directly mapped onto the ICD-10 coding system. We are the first to apply this methodology to find trajectories of chronic disease accumulation using this type of data.

2. Data

To perform the analysis we used the CPRD (*Clinical Practice Research Datalink*) dataset [3] that contains EHRs from general practices across the UK. Our study included 1.1 million English people aged 45 and over, followed up from 2001 to 2010. Since there is no clear definition of the chronic conditions that define multimorbid patients we combined 3 different approaches to decide which were the diseases of interest: a systematic literature review of the diseases included in multimorbidity studies, the NHS Quality and Outcomes Framework disease list and expert advice from primary-care clinicians and clinical epidemiologists. Our final list is shown in Table 1. The CALIBER platform [4] provided the data extracts and phenotyped the disease codes.

3. Methods

The method used is based on Jensen et al [2] and we adapted some aspects of the implementation to the specific circumstances of this analysis.

First, we computed the Relative Risk (RR) between all the pairs of diseases. The Relative Risk between disease A and B measures if A is a risk factor for B (i.e. a temporal correlation measure): $RR_{A \rightarrow B} = \frac{P(B_{t=T}=1|A_{t=0}=1)}{P(B_{t=T}=1|A_{t=0}=0)}$, where A_t and B_t are random variables of being diagnosed or not with diseases A and B at time t . We estimated 5-year RRs by matching each patient with $A_{t=0} = 1$ (exposed group) to N comparison groups where $A_{t=0} = 0$. The comparisons groups were matched at baseline by sex, age and deprivation index. We computed the P -values as the proportion of comparison groups where the count of cases with $B_{t=5} = 1$ is larger than in the exposed group. We selected all pairs $A \rightarrow B$ with $RR_{A \rightarrow B} > 1$ and P -value < 0.05 . In case of selecting both $A \rightarrow B$ and $B \rightarrow A$ we checked which direction was more significant. To evaluate the directionality we performed a Binomial test with the number of cases $N_{A \rightarrow B}$ and $N_{B \rightarrow A}$ in each direction. We kept at most one direction if it had P -value < 0.05 .

Once all pairs of relevant directions were obtained we went through the dataset to find patient trajectories formed only by these pairs. If a patient is diagnosed with $A \rightarrow B \rightarrow C \rightarrow D$ we considered that trajectory for our analysis if all the pairs $A \rightarrow B$, $B \rightarrow C$ and $C \rightarrow D$ are relevant. We only selected trajectories with at least 4 diagnosed conditions and with at least 5 patients going through that trajectory. Afterwards, we merged all these trajectories to form a network where each node represents a disease. Finally, we clustered the resulting network using the Markov Cluster Algorithm [5].

4. Results

We used RRs to measure the temporal correlation between the different diseases. To compute the RRs we used 1000 comparison groups for each exposed patient. In Figure 1 we show the significant 5-year RR between all pairs of diseases. From the 870 pairs of disease we find that 648 have $RR > 1$ and $P\text{-value} < 0.05$. There are 281 pairs of diseases (A, B) where $RR_{A \rightarrow B} > 1$ and $RR_{B \rightarrow A} > 1$ and the $P\text{-values}$ are under 0.05. Therefore, we checked for directionality to select the most significant directions of the two. After removing the less significant directed pairs we obtained 367 relevant directed pairs of diseases.

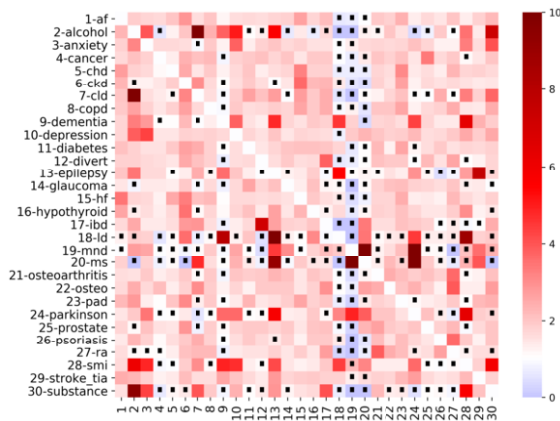


Figure 1. Relative Risks heatmap. The x-axis represents the baseline disease and the y-axis the follow-up disease. The x-axis and y-axis follow the same numerical index. The black dots indicate that $RR < 1$ or $P\text{-value} < 0.05$ so the relative risk is not significant. There are values with $RR > 10$ but for visualization purposes we have limited the colorbar range.

We searched disease trajectories from patients in the CPRD dataset formed exclusively by these significant pairs of diseases and with at least 4 diagnoses in the trajectory. Then, we merged all trajectories that have at least 5 patients going through them and we obtained a network of diseases where the diseases are the nodes and the edges are the transitions in the trajectories. We applied the Markov Clustering Algorithm [5] to the network and we obtain two clusters. The main cluster is shown in Figure 2 and contains 16 diseases. To visualize the second cluster in Figure 3 we added 3 disease long trajectories. There are some diseases that do not appear in any of the two clusters since they do not appear in 3 long patient trajectories formed by RR significant pairs. These

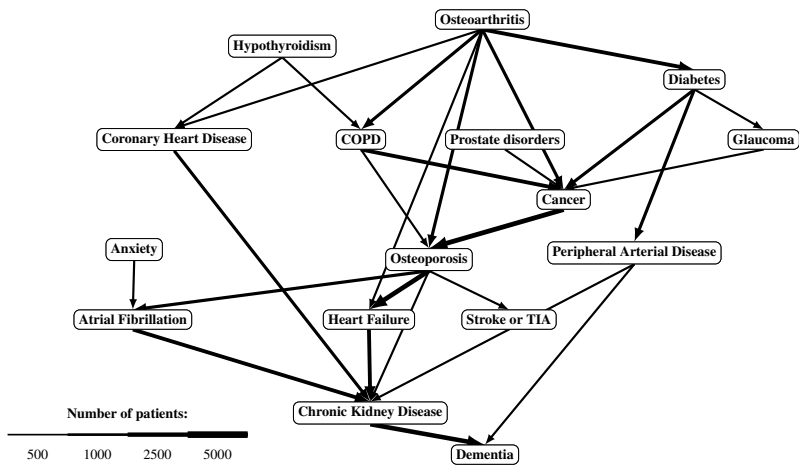


Figure 2. Main cluster of disease trajectories. The edge width represents the number of patients going through that diagnose sequence

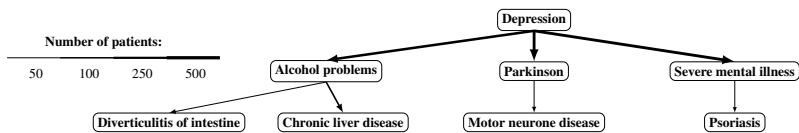


Figure 3. Second cluster of disease trajectories or Depression cluster. The edge width represents the number of patients going through that diagnose sequence

two clusters describe the most common trajectories of chronic disease accumulation. The main cluster network was build using trajectories of 4 conditions. Consequently, there are disease trajectories that start in a node with incoming edges.

5. Discussion

In this analysis we used EHRs from primary care to investigate the sequence in which patients get multiple chronic conditions. When studying multimorbidity it is not clear which are the chronic conditions that need to be taken into account. Thus, we decided to focus on a relatively small but prevalent number of diseases to understand what are the most common trajectories that multimorbid patients follow. We started by computing the RRs between diseases to evaluate the temporal correlation between any pair of conditions. Previous studies have focused on studying RRs of specific pairs of diseases. However, we were able to directly estimate 648 RRs thanks to the large dataset used. Most RRs related to learning disabilities, multiple sclerosis and motor neurone disease were found not significant due to the small amount of diagnosed cases. It is not possible to compare directly these RRs to previous studies because the time frame and disease definition can vary. Nevertheless, we found similar estimates for multiple pairs like CKD and Dementia [6] or PAD and CHD [7].

The main cluster shown in Figure 2 condenses the most common trajectories to multimorbidity into a single network. We can see how diseases like cancer, COPD, dia-

betes and osteoporosis are central in the progression. We also can observe clinically well known trajectories like *Diabetes* → *PAD* → *Stroke* [7]. The second cluster in Figure 3 shows reasonably expected alcohol problems related trajectories with depression as the initiating disease. However, there are many other trajectories that the two clusters do not cover. A dataset with a longer follow up period would help improve the results since we would be able to observe more complete trajectories. Previous work have applied this methodology to Secondary Care data and ICD-10 codes [2]. However, secondary care data does not capture most of the long lasting chronic conditions that multimorbid patients suffer from.

We know that sex and deprivation index are key factors to understand multimorbid patterns [8]. Consequently, in the future we should focus on finding specific trajectory networks for different population subgroups.

6. Conclusion

We used a data driven method to obtain networks that describe the main trajectories that patients follow to become multimorbid. The application of this method to a primary care dataset like CPRD confirms some of the already know temporal correlations between diseases and detects longer less known disease sequences. However, this method does not capture all the complex patterns that multimorbid patients follow. A main limitation of the model is that it only uses a directionality between the pairs of diseases. In the future we should focus on detecting specific trajectory patterns for different populations subgroups.

This work was partially supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care North Thames at Bart's Health NHS Trust. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The NIHR UCLH Biomedical Research Centre and EPSRC grant M020533/1 also supported this work.

References

- [1] MacMahon, S. Multimorbidity: A priority for global health research. *The Academy of Medical Sciences: London, UK* (2018).
- [2] Jensen, AB et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nature communications* **5** (2014), 4022.
- [3] Herrett, E et al. Data resource profile: clinical practice research datalink (CPRD). *International journal of epidemiology* **44** (2015), 827-836.
- [4] Denaxas SC et al. Data Resource Profile: Cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol* **41** (2012), 1625-1638.
- [5] Van Dongen, SM. Graph clustering by flow simulation. *Diss* (2000).
- [6] Helmer, C et al. Chronic kidney disease, cognitive decline, and incident dementia: the 3C Study. *Neurology* **77** (2011), 2043-2051.
- [7] Wattanakit, K et al. Risk factors for peripheral arterial disease incidence in persons with diabetes: the Atherosclerosis Risk in Communities (ARIC) Study. *Atherosclerosis* **180** (2005), 389-397.
- [8] Chan, MS. Socio-economic inequalities in life expectancy of older adults with and without multimorbidity: a record linkage study of 1.1 million people in England. *International journal of epidemiology* **48** (2019), 1340-1351.

Translating Social Determinants of Health into Standardized Clinical Entities

Michael WATKINS^{a,1}, Benjamin VIERNES^b, Viet NGUYEN^c, Leonardo ROJAS MEZARINA^{d,e}, Javier SILVA VALENCIA^e and Damian BORBOLLA^a

^aBiomedical Informatics, 421 Wakara Way, University of Utah, Salt Lake City, Utah 84108 ^bPopulation Health Sciences, 295 Chipeta Way, University of Utah, Salt Lake City, Utah 84108 ^cStratametrics LLC

^dInformation Systems Office, Jr. Capac Yupanqui 1400, National Institute of Health, Lima, Perú 15072

^eTelesalud Unit, Av Grau 755, Universidad Nacional Mayor de San Marcos, School of Medicine, Lima, Perú 15001

Abstract. Social determinants of health (SDH) are a valuable source of health information which still are not fully utilized in the clinical space. Knowing that a certain patient has trouble finding transportation, has a potentially hazardous relationship with a family member or close relative, is currently unemployed, or various other social factors would allow providers to tailor treatment plans in a way to best help that patient. However, these SDH must be gathered, represented, and stored in a standardized way before they can be leveraged by informatics tools designed for health providers. This process of translating SDH to standardized clinical entities includes two main steps. The first is a collaborative effort to establish an ontology of medical terminology codes (i.e., ICD, SNOMED, LOINC, etc.) which can be used to uniformly represent SDH as coded concepts. The second is a collaborative effort to use the FHIR standard to create profiles and extensions which will allow FHIR resources to be used to store the coded SDH as clinical entities. Each of these steps has their own complexities that must be considered and accounted for in future efforts to create interoperable clinical informatics solutions which utilize SDH.

Keywords. Social determinants of health, FHIR, interoperability, translational informatics

1. Introduction

The World Health Organization (WHO) recommends “systematically addressing” social determinants of health (SDH) to improve the health of the population as a whole [1]. Although there are many medical schools which provide training regarding SDH [2], some research has noted that current training strategies which “focus on knowing rather than acting cannot effect the change which will reduce health inequity in a meaningful way [3]. One recent study developed five categories of activities which should be used in tandem to fully address social needs. There, the authors describe the need of

¹ Michael Watkins, Corresponding author, Biomedical Informatics, University of Utah, 421 Wakara Way, Salt Lake City, Utah 84108, USA; E- mail: michael.watkins@utah.edu

developing an interoperable digital infrastructure that can easily share data between healthcare and social care organizations. The study also showed *awareness* as being just the first step in addressing a social issue; the other four categories include *adjustment*, *assistance*, *alignment*, and *advocacy* [4].

However, despite their potential implications in health, SDH are not a commonly utilized data source in most clinical workflows. Healthcare providers are already collecting some of this data, although this collection often occurs in an inconsistent and unstructured way. A structured, standardized method of SDH data collection by providers could inform individualized treatment, improve care coordination, and allow for a better population-level understanding of social and health equity for policy makers and healthcare system administrators. There are a few current informatics solutions that have sought to utilize SDH within the healthcare provider's office [5-6]. However, these are specific use-cases implemented among relatively small populations. In order to increase the population sizes included in such studies, interoperability must be a priority. This will allow cross-site deployment and wider-scale evaluation.

2. Background

The translation of SDH to standardized clinical entities is a complicated process due to the sheer number of stakeholders (i.e., government organizations, hospital systems, providers, clinical vocabulary creators, etc.) involved in the standardization of any clinical process. However, the informatics solutions discussed here will provide a high-level overview of the domains that decision-makers in SDH clinical implementation projects require.

The WHO defines SDH as “the conditions in which people are born, grow, work, live, and age, and the wider set of forces and systems shaping the conditions of daily life [1].” Marmot et al. summarized the relation of SDH to health as follows: “The poorest people have high levels of illness and premature mortality—but poor health is not confined to those who are worst off. At all levels of income, health and illness follow a social gradient: the lower the socioeconomic position, the worse the health [3].” Commonly assessed SDH include employment status, education level, food security, access to health services, housing status, income, discrimination, and social support [7].

The HL7 Fast Healthcare Interoperable Resources (FHIR) Specification is an ongoing effort to create a clinical data standard which enables the kind of interoperability which is necessary to fully utilize informatics tools and solutions. It abstracts clinical data concepts into separate resources. These resources are given unique identifiers which are used to string them together in order to represent all possible clinical scenarios [8].

The branch of translational informatics is becoming an increasingly important aspect of health informatics research. As its name suggests, translational informatics is the process of translating data from its original form to another using informatics methods. In health informatics, this has enabled various data sources to be utilized by informatics solutions in the clinical space in a standardized way. A sampling of these data sources includes environmental sensors, wearable health tools, imaging devices, various -omics databases, and detailed drug pharmaceutical research resources. In each case, the data is not formatted for clinical use and must be translated to a common clinical data standard before it can be used for interoperable clinical informatics solutions. Because these data are so potentially useful in clinical use, translational informatics has been named as an important milestone in achieving a nationwide learning health system [9].

3. Methods

The first step in a translational informatics approach to standardize the clinical use of SDH is to establish a shared ontology which coordinates the representation of any single determinant as a coded concept. This ontology could draw from established clinical coding systems such as SNOMED CT or ICD-10. Creating an ontology would allow separate efforts to build on each other rather than duplicate work.

One study worked to create a compendium of structured, codable social risk factors from common medical terminologies (i.e., SNOMED, LOINC, and ICD) [10]. Subsequent efforts should build upon such efforts in a collaborative way to achieve a robust common clinical data set (CCDS) of SDH which can be formally adopted and used for future SDH clinical solutions. Coding the determinants is the first step to various downstream interventions which enhance the value of care provided to patients. Standardizing the coding process with a shared ontology will result in a uniform knowledge source.

While mapping SDH to clinical codes is an important first step, it must be coupled with a standardized approach to storing them as clinical data entities. The clinical standard most suited for this purpose is the HL7 FHIR specification which is comprised of detailed resources (e.g. Observation, Condition, etc...). These resources require concepts to be coded to medical terminologies. However, each of these base resources would need to be properly extended in order to ensure that different research efforts would be able to maintain interoperability. This extension process relies on FHIR profiles to be created and published which specify how the relevant resource would need to be adapted to accurately represent the specific new data type in question. These efforts are typically undertaken by coordinated working groups and result in a FHIR Implementation Guide (IG). The IG walks future implementers through the process of creating the specific resources and extensions.

4. Results and Discussion

As is the case in any informatics efforts, collaboration and standardization are the catalysts for real progress. The creation of a robust ontology of SDH medical terminology codes, and the utilization of the FHIR standard for the standardized clinical representation of SDH, is a vast undertaking. The following considerations are additional points which should be carefully considered by those who join this effort.

A working group, known as the Gravity Project, has recently been announced which will tackle the complex issue of specifying FHIR profiles and extensions which can be used to store SDH [11]. Their coordinated efforts to profile various SDH will result in an IG to assist in the standardized adoption of FHIR resources to represent SDH. SDH are varied and many FHIR profiles may already exist for certain determinants in connection with other clinical use-cases. Care should be taken by researchers to ensure that existing work is built upon and only modified if necessary. FHIR can also provide important interoperability and standardization to the process of SDH collection. The FHIR Questionnaire resource is a versatile option for representing SDH-gathering questionnaires or question sets. It allows each question to be internally represented with a corresponding clinical code. SDH have historically been gathered through sets of questions which probe the patient's lifestyle and other aspects typically associated with SDH. Due to the nature of SDH data, its collection often requires patient self-reporting.

For example, although patient addresses are frequently used to infer socio-economic standing from census tracts, ZIP codes, and other community area designations, this method is dependent on survey and census material which is often outdated [12]. To assume that the patient is poor, based on that address alone, would result in incomplete information for that determinant. Since most approaches to clinical SDH gathering will use direct questions to avoid that unreliability, using FHIR Questionnaire resources would allow questionnaires to be used across institutions.

Responses to these questionnaires can be represented as FHIR QuestionnaireResponse resources. While using a FHIR Questionnaire ensures that the SDH are captured as coded concepts, an additional step is necessary to create a profiled FHIR resource for each relevant question response. For example, although a positive response to question on smoking history would result in a corresponding coded SDH within the QuestionnaireResponse, an additional step would be required to create a profiled FHIR Observation resource to represent that information. An informaticist querying for information on that patient's smoking history will be guided by the smoking status profile and will look for that corresponding resource, not for an obscure response within a QuestionnaireResponse resource. Coding a patient's smoking history as a FHIR Observation resource would allow for the data to be queried by relevant CDS solutions for all subsequent patient interactions. This method also enables patient-level, system-level, or population-level analysis of both the questionnaires and the unique, standardized clinical concepts.

The creation of an SDH ontology will be further complicated by the fact that there is a high level of overlap between the common medical terminologies. Reconciling this overlap will require collaboration and iterative feedback from relevant stakeholders. However, it is a crucial first step because despite the clinical standardization available through FHIR, the value is lost if the underlying representation of SDH is ambiguous.

5. Conclusion

Clinical informatics is a field which continues to present new opportunities for utilizing nontraditional medical data sources. Combining clinical informatics and translational informatics can be a powerful source of innovation. SDH represent a vast and varied source of data which can have important implications for health services. However, in order to use SDH in a clinical setting, they must be translated to standardized clinical entities. Following our approach will require a collaborative effort to curate a robust SDH ontology and to create the necessary FHIR profiles and extensions to store SDH in the clinical space. As this work moves forward, it will allow health professionals to provide more informed care for their patients and will likely inform later efforts to create new interoperable clinical informatics solutions which use other nontraditional data sources.

6. Acknowledgments

The work of this paper was supported in part by a World Health Organization WHO/PAHO, IS4H joint grant, between the University of Utah in the USA, and the National Institute of Health in Peru.

References

- [1] WHO | about social determinants of health. (2017). Retrieved from https://www.who.int/social_determinants/sdh_definition/en/
- [2] Sharma, M., Pinto, A., & Kumagal, A. (2018). From Making Lists to Making Change. Public Health Post, 1–3. Retrieved from <https://www.publichealthpost.org/research/teaching-social-determinants-of-health-making-change/>
- [3] Marmot, M., Friel, S., Bell, R., Houweling, T. A., Taylor, S., & Commission on Social Determinants of Health. (2008). Closing the gap in a generation: Health equity through action on the social determinants of health. The Lancet, 372(9650), 1661–1669. doi:10.1016/S0140-6736(08)61690-6
- [4] Bibbins-Domingo, K. (2019). Integrating social care into the delivery of health care. Jama, doi:10.1001/jama.2019.15603
- [5] Buitron de la Vega, Pablo, Losi, S., Sprague Martinez, L., Bovell-Ammon, A., Garg, A., James, T., . . . Kressin, N. R. (2019). Implementing an EHR-based screening and referral system to address social determinants of health in primary care. Medical Care, 57, S133–S139. doi:10.1097/MLR.0000000000001029
- [6] Bettencourt-Silva, J., Mulligan, N., Cullen, C., & Kotoulas, S. (2018). Bridging Clinical and Social Determinants of Health Using Unstructured Data. Studies in Health Technology and Informatics, 255, 70–74. <https://doi.org/10.3233/978-1-61499-921-8-70>
- [7] NCHHSTP Social Determinants of Health. (2014, March 10). Retrieved from <https://www.cdc.gov/nchhstp/socialdeterminants/faq.html>
- [8] FHIR Overview. (2018, December 27). Retrieved from <https://www.hl7.org/fhir/overview.html>
- [9] Friedman, C. P., Wong, A. K., & Blumenthal, D. (2010). Achieving a nationwide learning health system. Science Translational Medicine, 2(57), 57cm29. doi:10.1126/scitranslmed.3001456
- [10] Arons A, DeSilvey S, Fichtenberg C, Gottlieb L. (2018) Compendium of Medical Terminology Codes for Social Risk Factors. San Francisco, CA: Social Interventions Research and Evaluation Network; 2018. Available at: <http://sirennetwork.ucsf.edu/tools-resources/mmi/compendium-medical-terminology-codes-social-risk-factors>
- [11] Lovett, L. (2019, 21 August) *Social determinants of health program Gravity Project joins HL7 FHIR accelerator*. Retrieved from <https://www.mobihealthnews.com/news/north-america/social-determinants-health-program-gravity-project-joins-hl7-fhir-accelerator>
- [12] Shavers, V. L. (2007). Measurement of Socioeconomic Status in Health Disparities Research. Journal of the National Medical Association, 99(9), 1013–1023.

Trodden Lanes or New Paths: Ballisto- and Seismocardiography Till Now

Nico JÄHNE-RADEN¹, Henrike GÜTSCHLEG and Michael MARSCHOLLEK

*Peter L. Reichertz Institute for Medical Informatics
of TU Braunschweig and Hannover Medical School, Germany*

Abstract. Ballisto- and seismocardiography (BCG/SCG) are methods of studying blood circulation and cardiac function by using the vibration measurements at the body surface, e.g. via accelerometers. The aim of this work is to show the current relevance of BCG/SCG for the target medical diagnostics. To reach this goal and to examine the relevance, an overall search for all BCG and SCG articles in the databases PubMed and IEEEExplore was first carried out ("ballistocardiography OR seismocardiography") for the years till 2019. The results of this literature study show, overall 425 papers for the years from 2003 till 2019, with BCG (317) as significantly stronger represented than SCG (120). The distribution of the included subjects shows that a smaller group ($n \leq 10$) of mostly healthy people is more common. Last but not least, we examined which sensors have been included in the articles since 2003, with the result that accelerometers, whether as self-developed prototypes or installed in smartphones, were used in slightly less than 50% of the articles found. The differences in the numbers of publications between BCG and SCG may also be due to the distinction's complexity between BCG, which is more intuitive, and SCG. Looking at the number and distribution of included subjects, it is noticeable that this is rather low and primarily healthy subjects are used. However, the publication increase indicates that we are at a threshold in this topic and actual benefit to medicine.

Keywords. Review, Ballistocardiography, Seismocardiography

1. Introduction

Ballisto- and seismocardiography (BCG/SCG) are alternative methods for studying blood circulation and cardiac function. In this case, the BCG maps the pulse wave with its parameters via recording at the extremities, the head (often large vessels) or structures adjacent to the body [1,2]. The SCG is more likely to be recorded at the thorax and shows direct functional parameters of the heart, such as valve closure [1–3]. These methods can be used to perform long-term cardiovascular or respiratory illnesses, as a support or extension of common techniques such as the ECG.

As early as 1877 J. W. Gordon published a treatise on "Certain Molar Movements of the Human Body Produced by the Circulation of the Blood" and thus a cornerstone of the BCG [4]. Over the years, BCG and other forms, such as SCG, have evolved. New feature parameters such as pulse transit time as a marker for blood pressure, have come into focus [3,5]. But what about today with these methods of cardiovascular diagnostics.

¹ Corresponding Author, Nico Jähne-Raden, Hannover Medical School, Carl-Neuberg-Straße 1, D-30625 Hanover, Germany; E-mail: Nico.Jaehne-Raden@plri.de.

Especially related to the improvement of technical devices, e.g. sensors, in terms of size, cost or performance, BCG/SCG could contribute to improve diagnostics and care.

The motivation of this work is not only to show the current relevance of BCG/SCG for the target medical diagnostics, but also current research content, such as subject properties or measuring system setups. The aim is to be able to make a statement if the technology of BCG/SCG is really increasing and gaining importance in the research field of medical diagnostic.

This includes the questions about the origin of articles since 2003, as well as the investigation of different research circumstances, such as the extent of the subjects, the set of subjects, used sensors.

2. Method

To answer these questions from the introduction and to examine the relevance of BCG/SCG, an overall search for all BCG and SCG articles in the databases PubMed and IEEEExplore was first carried out ("ballistocardiography OR seismocardiography"). For the investigation of the articles found, after elimination of duplicates, no limitations were made on the date of publication or the settings, e.g. Animal experiments or similar.

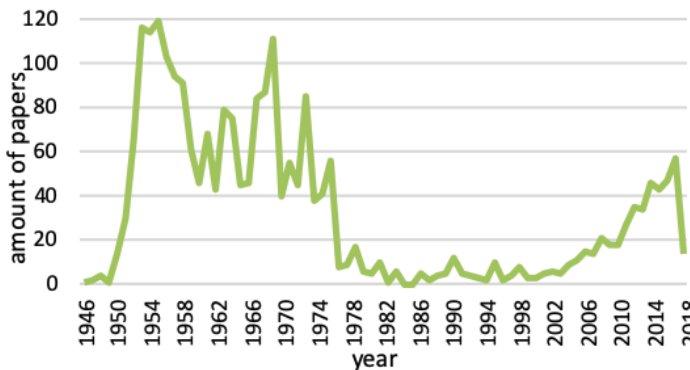


Figure 1. Amount of BCG/SCG papers over the years.

In Figure 1, the results of this overall search approach can be seen. We identified 2003 as the start of the renewed upswing in BCG/SCG research. With the adaptation of the literature search from May 2019 to the period beginning with the year 2003 up to the date of the search, the Prisma method for the literature search was continued [6].

After analyzing the paper found between 2003 and 2019, it was filtered for use in human medicine. In addition, review articles were excluded. With the inclusion of the found articles, the first author's seat, sensor technology, subject structure, number of subjects, general setting and the focus of the described work were extracted.

3. Results

The results of this literature study show, overall 425 papers over the years 2003 till 2019, that BCG is significantly stronger represented within the article than SCG (317/120).

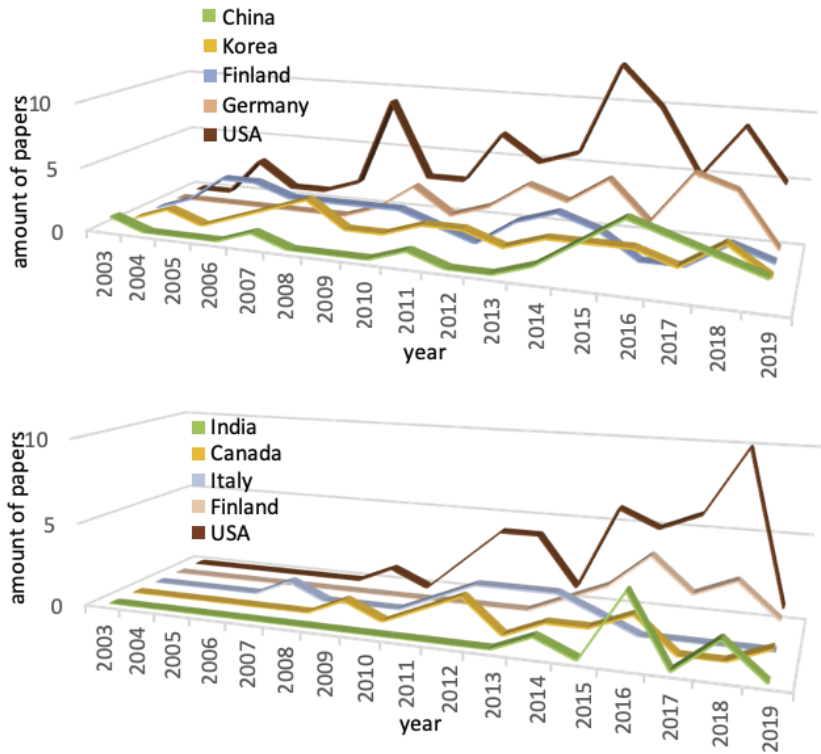


Figure 2. Amount of publication of BCG (first) and SCG (second) sorted by country (only top 5).

However, a slight increase after 2007 is to be noted, as can be seen in Figure 2, a rather linear, marked increase in publication figures overall. For the SCG, the number of releases increased since 2011. In both cases, the US is the country with the most publications over the years (see Figure 2).

In addition to the general results on the distribution of the articles found by years and countries, with regard to the focus on BCG or SCG, it was further considered how the subjects in the respective publications composed. This includes not only the number of subjects, but also the composition of the groups of subjects, for example by the inclusion of healthy subjects or subjects with heart diseases (e.g., heart failure) or other disorders (e.g., sleep disorders). In Figure 3, we graphically analyzed the results of the subject analysis with regard to the articles found. The number of subjects in Figure 3 (left) is the distribution of the subjects used, considering the number of found items. The figure clearly shows an increase in the number of 1 to 5, 10 and 20 subjects.

Looking now at the distribution of the subjects included in the articles found, Figure 3 (top left), a clear majority of the inclusion of healthy or heart-healthy subjects is evident. Last not least, we examined which measurement technologies have been included in the articles since 2003, with the result that accelerometers, whether as self-developed

prototypes or installed in smartphones, were used in slightly less than 50 percent of the articles found. The others were pressure sensors and weighing scales. Looking at the general focus of the articles found, the majority has a medical focus, such as the analysis of signal parameters for diagnostics (total of 338 articles). Only 72 articles have a more technical focus, such as the usability of specific sensors

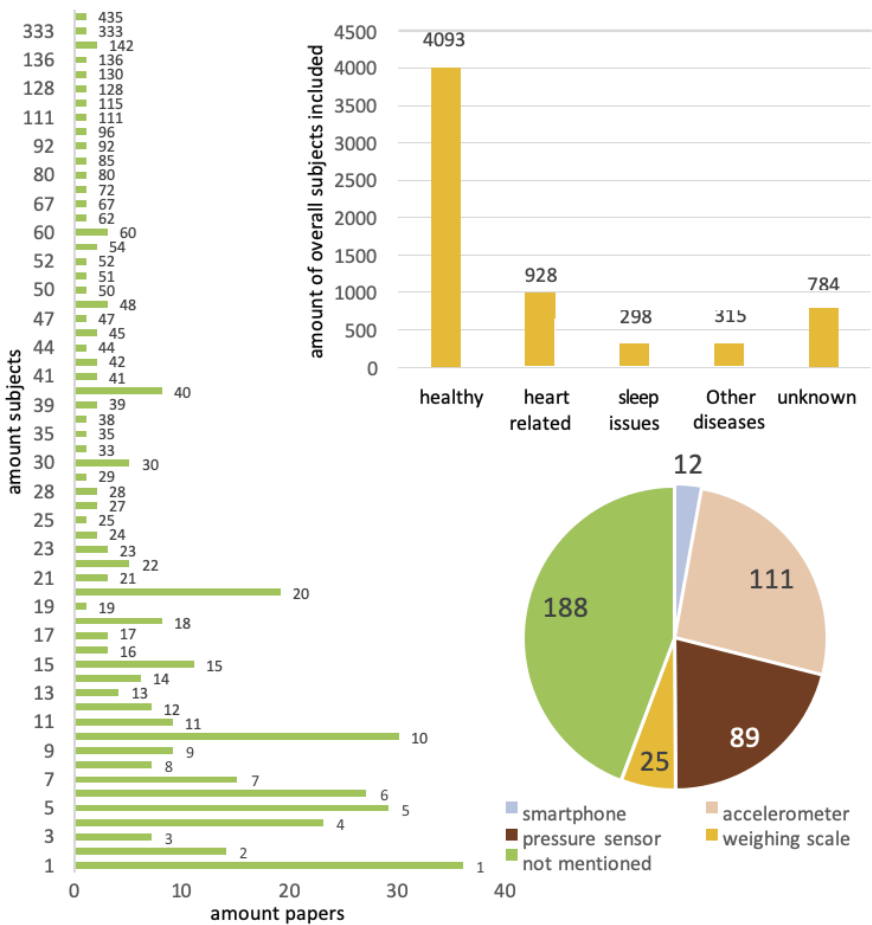


Figure 3. Number of used subjects in publications (left), type of used subjects (top right) and used sensing technology (bottom right).

4. Conclusion

Considering the listed results of this literature review on ballisto- and seismocardiography with a focus on the distribution to the focus area as well as the examination of the included subjects and the used sensor technology, an increasing number of publications starting from the year 2003 are shown. As a limitation of the review it has to be mentioned, that

we only used PubMed and IEEEExplore, which can be handled to mark a tendency and is not to be understood as complete. The publication slump which begins in the late 1970s can certainly be attributed to the development of image-based procedures, such as MRI and CT. The smaller peak in 2008 and the overall positive publication trend are certainly attributable to the smartphone as a new sensor system, as well as to technical developments (e.g., MEMS).

The differences in the numbers of publications between BCG and SCG may also be due to the complexity differences between the two methods. So, the BCG is a more intuitive method with less difficulty to identify features than the SCG. The increasing number of SCG publication figures can certainly be attributed to the increasingly sensitive sensor technology as well as the previous experience and findings in signal analysis. Moreover, Taebi et al. describe the challenges of the SCG, alone in the determination of the measuring position [7]. The distribution in the world shows the relevance of the topic worldwide, while North American, European and Asian research groups can be identified in the BCG area as well as in the SCG.

Looking now at the number and distribution of included subjects, it is noticeable that rather low and primarily healthy subjects are used. This is certainly due to the still early stage of BCG/SCG research. Of course, it is still kind of basic research, with meanwhile reaching consensus on basic signal BCG/SCG structures [see 2].

It is moreover interesting that beside the more obvious subject group of healthy and heart related subjects, the field of sleep issues are a noticeable field. For sure in a supine position while sleeping is near an optimal setup for BCG/SCG measurements.

But once again, SCG complexity shows itself (see also [7]). However, the increase in "diseased" subjects and in the number of papers in the recent years indicates that we are at a threshold in this topic and the actual benefit to the medicine, e.g. as diagnostic support, may be in the near future.

References

- [1] L. Giovannardi, O.T. Inan, R.M. Wiard, M. Etemadi, and G.T.A. Kovacs, Ballistocardiography--a method worth revisiting. *Conf Proc IEEE Eng Med Biol Soc* 2011 (2011), 4279–4282.
- [2] O.T. Inan, P.-F. Migeotte, K.-S. Park, M. Etemadi, K. Tavakolian, R. Casanella,
- [3] J. Zanetti, J. Tank, I. Funtova, G.K. Prisk, and M. Di Rienzo, Ballistocardiography and seismocardiography: a review of recent advances. *IEEE J Biomed Health Inform* 19 (2015), 1414–1427.
- [4] J.M. Zanetti and K. Tavakolian, Seismocardiography: past, present and future. *Conf Proc IEEE Eng Med Biol Soc* 2013 (2013), 7004–7007.
- [5] J.W. Gordon, Certain Molar Movements of the Human Body produced by the Circulation of the Blood. *J Anat Physiol* 11 (1877), 533–536.
- [6] S.L.-O. Martin, A.M. Carek, C.-S. Kim, H. Ashouri, O.T. Inan, J.-O. Hahn, and
- [7] R. Mukkamala, Author Correction: Weighing Scale-Based Pulse Transit Time is a Superior Marker of Blood Pressure than Conventional Pulse Arrival Time. *Sci Rep* 8 (2018), 15838.
- [8] A.A. Selçuk, A Guide for Systematic Reviews: PRISMA. *Türk Arch Otorhinolaryngol* 57 (2019), 57–58.
- [9] A. Taebi, B. Solar, A. Bomar, R. Sandler, and H. Mansy, Recent Advances in Seismocardiography. *Vibration* 2 (2019), 64–86.

Understanding Petri Nets in Health Sciences Education: The Health Issue Network Perspective

Fabrizio L. RICCI^{1a}, Fabrizio CONSORTI^{b,c}, Fabrizio PECORARO^a, Daniela LUZI^a, Valentina MINGARELLI^b, Sara MIOTTI^b and Oscar TAMBURIS^{a,d}

^a *Institute for Research on Population and Social Policies, National Research Council, Rome, Italy*

^b *Dept. of Surgical Sciences, University “Sapienza”, Rome, Italy*

^c *Italian Society of Medical Education (SIPEM), Italy*

^d *Dept. of Veterinary Medicine and Animal Productions, University “Federico II”, Naples, Italy*

Abstract. Scarce literature exists as to the use of Petri Nets (PN) to model the dynamic evolution of health issues in a deterministic way. Starting from the HIN (Health Issue Network) approach, the paper aims at describing the suitability of PN in supporting the Case-Based Learning method for improving an educational simulation environment in which students can manage realistic clinical data related to the evolution of a patient's health state over time.

Keywords. Health Issue Network, Case-Based Learning, Electronic Health Record, Petri Nets, Medical education

1. Introduction

The increase of both life expectancy and incidence of chronic diseases and disabilities in western societies that goes under the name of “epidemiological transition” [1] implies that nowadays physicians must provide healthcare activities over time to patients with multiple chronic diseases and, eventually, with an acute concomitant condition. Unfortunately, the concept of time is not present in current models of clinical reasoning [2-4], as they focus on the cognitive process of making a diagnosis and do not address the problem of how medical conditions evolve over time in a patient's clinical history, and/or how each condition interferes with the evolution of the other co-existing conditions. Case-Based Learning (CBL) is one of the most effective methods for teaching/learning clinical reasoning [5]. CBL has been successfully implemented through computer-based simulations [6], but these “virtual patients” are generally acute cases rarely described under the perspective of health evolution over time. Although the clinical question “What is the diagnosis in this patient?” remains pivotal, our aim was to develop the ability to answer to three other important questions, that is: starting from the available information on a patient in his/her present state (i.e.

¹ Fabrizio L. Ricci, IRPPS CNR, via Palestro, 32 – 00185 – Rome, Italy; E-mail: f.ricci@irpps.cnr.it

given a specific pathophysiology and treatment), (1) how health conditions may evolve in the future given the present patient's health status? (2) which past health conditions have possibly evolved in the present health status? (3) which health conditions have influenced the evolution of other health issues?

The HIN (Health Issue Network) approach was introduced [7] to improve CBL in a simulation environment for medical education. HIN allows: (i) reconstructing the implicit knowledge that lies behind and origins the physician's way of thinking, turning it into an explicit knowledge that makes possible to extract real clinical cases from Electronic Health Records; (ii) formalizing the description of an evolution path of a health issue of a person, through the implementation of the Petri Nets (PN) formalism [8], which is capable of translating HIN's key concepts in a graphical language. In HIN the PN marking represents the health status of a person, which evolves from one issue to another: each place node can only contain one token at a time. The transition from a HI (place node of the PN) occurs via a well-defined evolution (transition node of the PN) – PNs are bipartite graphs. Moreover, HIN model is a P/T safe net [8]. More details on the HIN model can be found in [7].

The present work investigates more in depth how Petri Nets, and therefore the HIN model, are an accurate way to represent the evolution of a patient's health state over time. The article summarizes the main features of the HIN model, discusses why PNs are suitable as a formal model underlying the HIN model and highlights the added value of HIN project.

2. Basic concepts of Health Issue Network

The guiding idea of HIN is a network of clinical problems as a basis for modeling a person's medical history. Each episode of care starts with an initial clinical problem (e.g. a symptom) and is described by its evolution together with the healthcare activities that make up the care path implemented to solve the original issue. The HIN model is characterized by two key concepts underlying the evolution of a health state:

(i) *health issue* (HI), that is any problem related to a patient's health (diagnosis, diagnostic hypothesis, symptom, sign, condition or risk factor, iatrogenic problem, general class of problems) [9], and; (ii) *evolution*, that is any change in a HI through which the patient moves from one health state to another. Five fundamental evolutions have been identified: *examining in-depth*, *worsening*, *complication*, *recurrence* and *improvement*. A HI can evolve for its own dynamics and/or – if more than one HIs occur in the same patient – be influenced by other HIs in its evolution. Both evolutions and influences are intended as probabilistic events, but in the real history of a patient, a HI evolution has occurred or not: alternative paths (evolutions in conflict) do not exist.

3. The choice of Petri Nets

In order to represent the evolution path of a person's health state, the choice of the Petri Nets formalism is based on the characteristics of the evolution path, as these well match with the properties of the PNs. Moreover, PNs allow to reduce the complexity of real clinical cases of patients with chronic multimorbidity into a rather simple mathematical model, which allows a representation of cases suitable for an educational environment, in particular: realistic as to its ability to reproduce the concepts of

evolution of HIs and influence between HIs; customizable according different learning outcomes. Here following are some of the properties of PN related to their use in a HIN-based simulation environment to represent clinically meaningful situations:

the PN is a *linear system* – the aim of the HIN approach is to extract real cases from medical records and let the students carry out exercises based on real/realistic data. The learning outcome is the ability to identify the HI evolutions that really occurred. A patient's clinical history is the “sum” of the evolutions of each HI and, at any given time, the patient’s health state is the “set” of the active HIs. The evolution over time of a patient's health state can then be represented as a linear system composed by a series of defined states. For example: in a patient with chronic pulmonary disease (COPD) and initial prostatic hypertrophy (BPH) these co-existing HIs make up the initial health state and the clinical history is composed by all the evolutions of these HIs. PNs represent a series of transitions in a deterministic manner. Moreover, in PNs the effect of a “sum” of input perturbations is equal to the “sum” of the effects produced by each individual perturbation, with the possibility of breaking up a linear problem;

the PN is a *discrete distributed system* – not all HIs are connected to each other: this implies that the evolution of a patient’s health state actually comprises parts (subnets) that can be independent from each other. In the above example, the BPH can worsen and produce the need for surgery without any influence on COPD. A PN – as well as the system of the evolving health states;

the PN (as evolution of a patient's health state) is an *asynchronous system* – the evolution concerns one HI at a time and not more HIs together. Since each health state can be interpreted as sum of several partial and independent subnets, an evolution only influences a part of the overall state; the firing of an evolution at a time preserves the locality of evolution of HIN diagram: two unrelated evolutions “never occur at the same time”. In the above example, surgery makes the HI BPH evolve into “operated BPH”. Because of the process of care (surgery) the HI COPD may evolve in “acute respiratory failure”. If the patient has another unrelated HI (for example a symptomatic uncomplicated diverticular disease), this can evolve or not, but in any case, its evolution is recorded in a different time than the other ones;

a PN is a *system without memory* – when an evolution takes place, a new health state must be assessed to envisage potential health evolutions: new possible evolutions may have been enabled and some others disabled; the identification of the new potential evolutions doesn’t depend on how the new health state has been reached. Furthermore, for the educational purposes of HIN, the input to every single evolution of an HI is made up of all the HIs that make the firing of that evolution possible. In the above example, in managing the patient with a present state of BPH and COPD, a physician can only envisage through what series of evolutions the patient got to his present state. The set of previous evolutions are summarized in the actual state but cannot be detailed or inferred in any way.

Table 1. The correspondence between evolution for a patient's health state and for Petri Nets

Evolution of a patient's health state	Petri Net
Health Issue	Place
Evolution	Transition
Evolution path of a Health issue	Firing sequence
Health state	Marking
Clinical history	Reachability graph

Tab.1 shows the kernel of the correspondence between the concepts of the evolution for a patient’s health state and those for a PN system.

4. Results: the use of Petri net in the simulation environment

The HIN method also deploys the computational properties of PN_s. The three previously reported clinical questions can be solved in terms of the PN property of reachability [8]. These questions can be expressed as: (i) which HIs are achieved starting from a specific health state? (ii) which HIs are needed to reach a specific HI? (iii) what link exists between two identified HIs? Example: the learning outcome is the management of a patient with chronic co-morbidities (cardiac and renal failure) undergoing abdominal surgery for cancer and developing complications. The left box of fig. 1 shows the PN that represents the prototypical clinical condition, while the right box shows the HIN of a real case, selected from a database of medical record according to the selection criteria derived from the learning outcome. The real case has much more details and possibly slightly different paths of evolution.

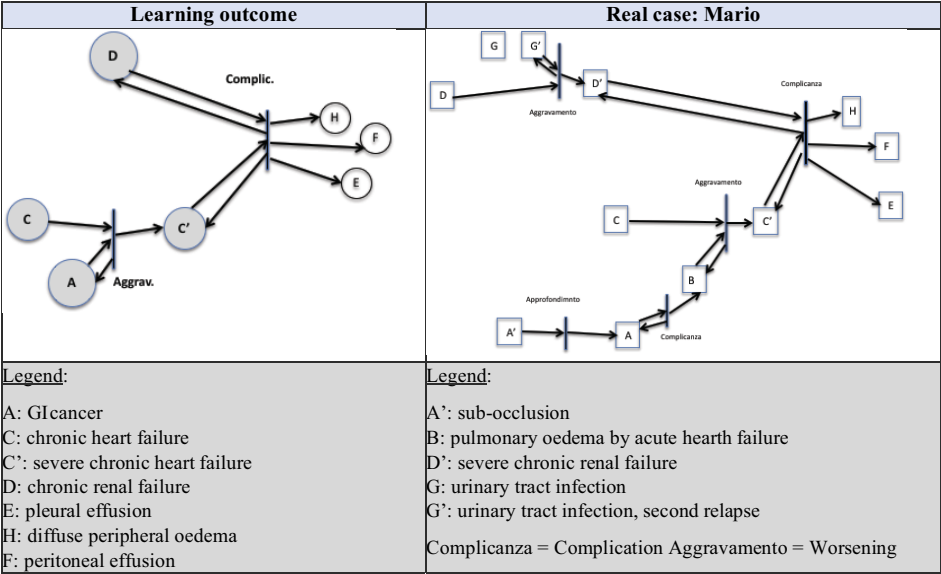


Figure 1. An example of HIN diagrams

Examples of the above questions (right box of fig. 1) are the following: (i) which HIs are achieved starting from the health state {D', C, A'?} (ii) which HIs are needed to reach the HI C'? (iii) what link exists between the HIs C' and A'? In modelling a health state as a set of HIs, and therefore as a marking in the HIN model, PN_s allow solving these clinical questions by means of the reachability graph and of a firing sequence. Another use of PN_s is to measure the “distance²” between two HIN diagrams. This problem is solved via a rewriting of a PN (e.g. deleting a place-HI or a transition-

² We indicate with the term *distance* the difference in terms of places, transitions and edges, besides the possibility that the two diagrams can become equal after their rewriting.

evolution, or an edge). The analysis of graph distance is useful in the following scenarios:

- Choice of the real case – how far is the HIN diagram of a real case from the HIN diagram of the condition in the learning outcome? In the example of Fig. 1 the “distance” was analyzed between the HIN diagram for a didactic goal, and the HIN diagram of a real case (Mario).
- Given a set of HIs (without the related evolutions) with an indication of the initial and final health states, the learner is asked to envisage the possible evolutions of the health state, i.e. to draw the evolutions and influences and link them to the related HIs: how different will be the resulting HIN diagram drawn by the learner from the HIN diagram drawn by the teacher as correct solution of the exercise?

5. Discussion and Conclusions

Scarce literature exists as to the use of modelling tools for representing the evolution of a health issue. A number of papers describe modelling tools for developing workflows [10], protocols [11], or guidelines [12], but in all these cases the focus is on a flow of clinical activities planned and executed so as to accomplish specific tasks, or to work out clinical issues. No mention is made for what concerns the way activities can change the health state, or how this evolves over time. From an educational point of view as well, few works have been produced to develop educational models and tools for the competence of clinical management of chronic and multimorbidity conditions. Basically, the indications are about practical training and the use of virtual cases, which are not based on a sound theoretical model [13]. The HIN project is developing an integrated environment of theoretically based technical solutions to design, implement and assess the simulation of clinical cases of chronic and multimorbidity patients [7].

References

- [1] Omran AR. The epidemiologic transition: a theory of the epidemiology of population change. 1971. *Milbank Q.* 2005;83(4):731-57.
- [2] Croskerry P. *A universal model of diagnostic reasoning.* *Acad Med.* 2009;84(8):1022-8.
- [3] Stoeckle JD. The market pushes education from ward to office, from acute to chronic illness and prevention: will case method teaching-learning change? *Arch Intern Med.* 2000;160(3):273-80.
- [4] Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ.* 2005;39(1):98-106.
- [5] Thistlethwaite JE, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Med Teach.* 2012;34(6):e421-44.
- [6] Consorti F, et al. Efficacy of virtual patients in medical education: A meta-analysis of randomized studies. *Computers & Education.* 2012; 59: 1001–1008
- [7] Ricci FL, et al. HIN -Health Issue Network as Means to Improve Case-Based Learning in Health Sciences Education. *Stud Health Technol Inform.* 2018; 255:262-266.
- [8] Peterson JL. *Petri Net Theory and the Modeling of Systems.* Prentice Hall, ISBN 0-13-661983-5, 1981.
- [9] ContSys. *A system of concepts for the continuity of care.* Available at www.contsys.org
- [10] Fanti MP, et al. *A Petri Net Model of an Integrated System for the Health Care at Home Management.* 2014 IEEE International Conference on Automation Science and Engineering (CASE), 2014.
- [11] Whittaker SJ. *Augmenting Petri Nets to Model Health-Care Protocols.* Ph.D. Dissertation. Queen's University, Kingston, Ontario, Canada. (AAINR78474), 2011.
- [12] Beccuti M, et al. *Modeling clinical guidelines through Petri Nets.* Conference on Artificial Intelligence in Medicine in Europe, 2009.
- [13] Khin-Htun S, Kushairi A. Twelve Tips for Developing Clinical Reasoning Skills in the Pre-Clinical and Clinical Stages of Medical School. *Med Teach.* 2019 Sep;41(9):1007-1011.

Using Big Data Analytics to Identify Dentists with Frequent Future Malpractice Claims

Wanting CUI^{1a}, Joseph FINKELSTEIN^a

^a*Icahn School of Medicine at Mount Sinai, New York, NY, USA*

Abstract. Healthcare spending has been growing at an increasing rate in the US, due in part to medical malpractice costs. Dental malpractice is an area that has not been studied in depth. Using National Practitioner Data Bank (NPDB), we explored the extent of dental malpractice claims and sought to construct a predictive model that can help us identify dental practitioners at risk of performing medical malpractice. Over 1,500 dental malpractice claims were reported annually, and over \$1.7 billion being paid out by medical malpractice insurers over the past 15 years. Majority of claims resulted in minor injuries, and the number of major injury claims increased over years. In prediction, we randomly split the data into train (75%) and test (25%) datasets. We trained and tuned models using 5-fold cross validation on the training set. Then, we fitted the model on the test data for performance measures. We used Logistic Regression, Random Forest (RF) and XGBoost and tuned the hyperparameters of models accordingly through grid search and cross validation. XGBoost was the best machine learning model to predict the risk of dentists having several malpractice reports. The best performing model had an accuracy of 72.8% with 30.6% F1 score. The NPDB database is a valuable dataset to study dental malpractice claims. Further analysis of information extracted from this dataset is warranted.

Keywords. Data Science, Big Data Analytics, Machine Learning, Predictive Model

1. Introduction

The United States currently ranks highest in health care spending among the developed nations of the world. In 2017, health care spending was \$3.5 trillion in total, which accounted for a 17.9% share of the nation's GDP. It has also increased over 50% in the past 10 years, compared to \$2.3 trillion in 2007 [1]. One of the reasons for the increasing cost of health care is defensive medicine. Defensive medicine refers to physicians prescribe diagnostic test or medical treatment that depart from normal medical practice as a safeguard from litigation [2]. According to a recent study, there is a statistically significant correlation between specialists concerns regarding potential medico-legal disputes and the choice of defensive medical procedures [3].

Although the number of malpractice claims has decreased since 2001, they are still very common in the United States. According to a 2016 Benchmark Survey, a third of

¹ Wanting Cui, Icahn School of Medicine at Mount Sinai, 1770 Madison Ave, New York, NY, USA, 10035, E-mail: wanting.cui@mssm.edu

physicians have been sued at least once in their career [4]. In addition, the average costs of malpractice cases have increased year over year. The malpractice insurance premiums have also increased accordingly, due to the high cost of malpractice claims. As a result, physicians' incomes may be affected by the rising cost of doing business as well.

While physicians' (MD, DO) malpractice cases have been frequently studied in recent years, there is little study on dental malpractice disputes and predictive models on future prevention are rarely constructed [5, 6]. In this article, using data from the National Practitioner Data Bank, we reported findings concerning the distribution of dental malpractices and the extent of dental malpractice risk. By studying malpractice, we hope to mitigate the high insurance premiums caused by the prevalence of malpractice claims. Our objectives were to review recent trends in dental malpractices, based on cost, number of claims, severity of injuries, physicians' and plaintiffs' information; and to create machine learning models predicting dentists with higher risk of frequent malpractices based on their first malpractice claim.

2. Method

2.1. Dataset

The dataset comes from National Practitioner Data Bank's (NPDB) Public Used Data File [7]. National Practitioner Data Bank, started in 1990, is a database that contains information related to the professional competence and conduct of physicians, dentists, and other health care practitioners. We accessed the data in 2019, so it included reports received from September 1, 1990 through December 31, 2018.

This analysis studied dental malpractice reports aiming to find important factors and to identify possible practitioners who could have repeat acts of malpractice in the future based on characteristics of their first offence. Due to several new variables added since January 31, 2004, and an average 4-year-gap between an incident and its payment, a 10-year-period (2005 – 2014), based on the report filing year (ORIGYEAR) was used in this study. In addition, only dentists (LICNFELD: 30) and dental residents (LICNFELD: 35) were included in this study. Records with missing data were deleted. Furthermore, only the first malpractice report of each practitioner was kept.

2.2. Variable Selection

Columns that are irrelevant to malpractice, such as Adverse Action Classification, Basis for Action, etc., were deleted from the dataset. After adjusting for inflation based on CPI-U for the U.S. City Average for All Items from the Bureau of Labor Statistics, we replaced Payment (PAYMENT) and Total Payment (TOTALPMT) by PAYMENT_ADJ and TOTALPMT_ADJ correspondingly. According to guidelines, we created a new variable 'STATE', which equals Work State if a work state value was reported and Home State if no work state was reported [7]. Since, for most practitioners, their License States (LICNSTAT) and their work States (STATE) are identical, we created a new binary variable State Difference (STATEDIFF), which is 1 if license state is different from state, and 0 otherwise. We counted the occurrences of each LICNSTAT, weighed against population of that state and rescaled those values to between 0.0001 and 0.9999. This information was stored in variable

'STATWEGHT_ADJ'. Similar to States, we created a new variable Payment Difference (PAYDIFF) to track the difference between payment and total payment. In addition, we created a variable Years of Experience (YREXP), for years between Graduation (GRAD) and Year of Act or Omission 1 (MALYEAR1). We also added variable Duration for years between MALYEAR1 and Year of Original Report processed (ORIGYEAR). A binary prediction variable 'RISK' was created based on the number of malpractice reports a practitioner had. The risk is 0 for a practitioner with only one malpractice report on record and the risk is 1 if a practitioner has more than one malpractice reports.

2.3. Statistical Methods

In variable analysis, we calculated summary statistics, such as mean, median and quantiles of continuous variables using standard methods. For those variables with skewed distribution, the Mann-Whitney test was performed to compare the difference between two risk levels. For variables with normal distribution, a two sample t-test was used. In addition, Pearson's Chi-Square test was used to identify differences between some categorical variables against the two risk levels.

In prediction, we randomly split the data into train (75%) and test (25%) datasets. We trained and tuned models using 5-fold cross validation on the training set. Then, we fitted the model on the test data for performance measures. We used Logistic Regression, Random Forest (RF) and XGBoost and tuned the hyperparameters of models accordingly through grid search and cross validation. Due to imbalanced data, we experimented with up sampling, down sampling and Synthetic Minority Over-sampling Technique (SMOTE).

All analyses were performed in RStudio (R version 3.5.3). All tests were two-sided, with $p < 0.05$ being considered statistically significant.

3. Results

3.1. Variable Analysis

Between 2005 and 2014, there were over 9,000 new dental practitioners who had their first malpractice cases reported. Of those practitioners, 7,494 (83.22%) of them were identified as risk free for multiple malpractices in the near future (Risk 0), and 1,511 (16.78%) of them were identified as at risk of multiple malpractices in the near future (Risk 1). Improper performance was the primary type of complaint.

Despite some fluctuations, the total number of malpractice reports generally decreased from 2005 to 2014. There were over 1,000 cases in 2005; this decreased to 857 cases in 2014 (Figure 1A). Although the number of cases decreased throughout the ten years being studied, mean payment amount after adjusting for inflation (PAYMENT_ADJ) increased for both risk free and at risk practitioners. This suggests that the cost of malpractice increased every year (Figure 1B). In addition, mean payment amount (Risk 0 = \$70,360; Risk 1 = \$105,660) and median payment amount (Risk 0 = \$22,436; Risk 1 = \$29,570) for at risk cases were higher than those for risk free cases (p value < 0.001). Average years of experience (YREXP) for the risk free group (mean = 24.17) was also significantly higher ($P < 0.001$) than that of the at risk group (mean = 22.96). Most practitioners have their first claim in their 40s.

Furthermore, there was a significant difference in the duration (DURATION) of each dispute between the two groups (p value < 0.001). The average duration of the risk free group was 3.27 years, in contrast to 3.93 years of the at risk group.

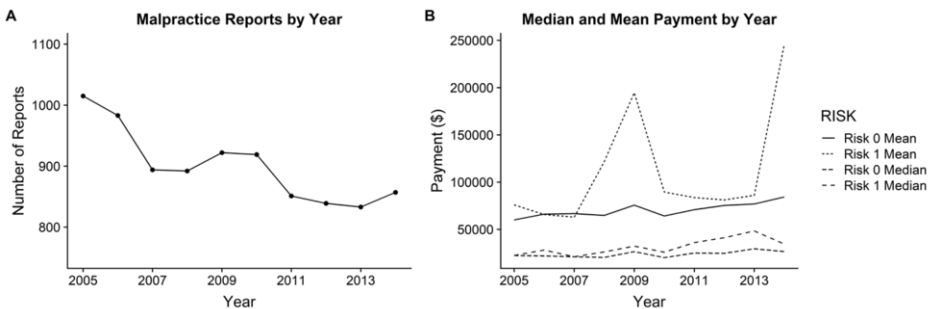


Figure 1. Malpractice reports trend.

There was no significant difference for different levels of outcome (OUTCOME) between the two groups (p value = 0.1159). While the amount of minor injuries decreased gradually from 2005 to 2014, the amount of major injuries increased. The amount of emotional injury and death remained constant. Moreover, there was no significant difference for gender of patients (PTGENDER) between groups (p value = 0.4). However, around two-third of the reports involved female patients and the proportion stayed consistent over the years.

3.2. Predictive Models

We used Logistic Regression as a baseline model. Prior to resampling the data, the logistic model categorized a significant amount of data to the risk free group, due to the imbalanced dataset. Thus, we experimented with up sampling, down sampling and SMOTE and compared the results. The result using Logistic Regression yielded a higher F1 score after resampling, increasing from 2.1% to 30.6%. However, the accuracy of the model decreased from over 80% to 60%. Thus, we would like to find a model that produced a high F1 score without the cost of accuracy. Both Logistic Regression and XGBoost produced relatively high F1 score. Up and down sampling methods also yielded better results than the SMOTE method (Table 1).

Considering XGBoost has a faster model training time and is a robust model which is less subject to overfitting, we decided to use XGBoost as our primary model. We fine-tuned hyper parameters of XGBoost using grid search. We also combined both up sampling and down sampling techniques using the ROSE package in an attempt to find balance between these two methods. The final model accuracy was 72.8% and F1 score was 30.6%.

Table 1. Predictive models results

Logistic Baseline	No Sampling (83.7%, 2.13%)		
	Down (Accu, F1)	Up (Accu, F1)	SMOTE (Accu, F1)
Logistic	(59.1%, 30.7%)	(60.6%, 30.9%)	(74.3%, 24.5%)
Random Forest	(57.7%, 30.6%)	(81.7%, 13.8%)	(81.4%, 11.8%)
XGBoost	(58.5%, 31.0%)	(61.6%, 30.0%)	(82.8%, 10.6%)
XGBoost (Best)	Up and Down Sampling (72.8%, 30.6%)		

4. Discussion

The final predictive model supports the previous discovery that payment amount, dentists' years of experience and length of disputes are major variables in determining whether a dentist is at risk of multiple malpractice cases in the near future. In addition, allegation type, dentists' working states and patient ages are also important factors for prediction.

A highly imbalanced dataset and categorical variables with multiple levels are the two main challenges for constructing predictive models. We employed different sampling techniques and used robust models that are suitable for categorical data to solve these issues. In future studies, we plan to optimize the probability threshold to overcome the imbalanced class problem. Furthermore, summary information of the dataset will be passed into the XGBoost model as additional features to improve model accuracy and F1 score.

5. Conclusion

In the past 15 years, over 1.7 Billion dollars of dental malpractice payments were awarded to patients. While the number of malpractices claims decreased through years, the payment amount per claim has increased. Practitioners' age was also an important variable. Dentists with more years of experience and older age were subjected to higher risk of having multiple malpractice cases in the near future. Additional standardized training should be provided to overcome improper performance issue.

XGBoost was the best machine learning model to predict the risk of dentists having several malpractice reports. The best performing model had an accuracy of 72.8% with 30.6% F1 score.

The NPDB database is a valuable dataset to study dental malpractice claims. Further analysis of information extracted from this dataset is warranted.

References

- [1] National Health Expenditure Data: Historical, U.S. Centers for Medicare & Medicaid Services, CMS.gov.
- [2] Sekhar, Msonal, and N. Vyas, Defensive Medicine: A Bane to Healthcare, *Annals of Medical and Health Sciences Research* **3**(2) (2013), 295.
- [3] Motta S., Testa D., Cesari U., Quaremba G., & Motta G., Medical Liability, Defensive Medicine and Professional Insurance in Otolaryngology, *BMC Research Notes* **8** (2015), 343.
- [4] Guardado J, Policy Research Perspectives: Medical Liability Claim Frequency among U.S. Physicians, American Medical Association.
- [5] Saber Tehrani, A. S., Lee, H., Mathews, S. C., Shore, A., Makary, M. A., Pronovost, P. J., & Newman-Toker, D. E., 25 Year Summary of US Malpractice Claims for Diagnostic Errors 1986–2010: An Analysis from the National Practitioner Data Bank, *BMJ Quality & Safety* **22**(8) (2013), 672– 680.
- [6] Rubin J.B., Bishop T.F., Characteristics of Paid Malpractice Claims Settled In and Out of Court in the USA: a Retrospective Analysis, *BMJ Open* **3**(6) (2013), e002985.
- [7] National Practitioner Data Bank Public Use Data File, 2018, U.S. Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Workforce, Division of Practitioner Data Bank.

Using Business Intelligence Tools to Support Medical Validation of Laboratory Tests

Gian-Andrea DEGEN^{a,b}, Viola GÜNTHER, Jürgen HOLM^a, Thomas BÜRKLE^a, and Murat SARIYAR^{a,1}

^a*Bern University of Appl. Sciences, Dept. Medical Informatics, Switzerland*

^b*Institute for Clinical Chemistry, University Hospital Zurich, Switzerland*

Abstract. Modern clinical laboratories have to confirm that the procedures used for specific tests are reliable and valid. There are several sources of errors and interferences that can invalidate the results. Medical validation refers to the plausibility check of the test results. Implausible results indicate that something might went wrong with a sample retrieved from the patient, e.g., the blood sample got contaminated with another fluid, which requires re-examination. Here, we describe how an integrated R-based business intelligence (BI) tool can be developed that increase the efficiency of the medical validation at the Institute of Clinical Chemistry (ICC) of the University Hospital Zurich. A BI software environment allowed us to digitalize steps in the validation process that were manually done in Excel worksheets, e.g., importing the data, calculating percentiles, and producing graphical outputs.

Keywords. Clinical validation; delta check; extreme value analysis; business intelligence; R

1. Introduction

It is indispensable for modern clinical laboratories to confirm that the procedures and methods used for specific tests are reliable and valid. There are several sources of errors and interferences that can invalidate the results. One important distinction in this regard is between technical and medical sources [1,2]. The former one relates to the quality of laboratory devices and instruments used, e.g., purity of chemicals or adequacy of materials. Whenever a test is introduced, changed or used in a new context, technical validation should be triggered. There are several guidelines and quality standards for ensuring technical validity [3]. Medical validation, on the other hand, refers to the plausibility of the test results. Assuming technical validity, implausible results indicate that something went wrong with a sample retrieved from the patient, e.g., a blood sample was contaminated with another fluid, which requires re-examination.

Medical validation consists of two parts: transversal and longitudinal assessments, addressing the plausibility of results with respect to a population or to an individual [4]. Transversal assessments can be based on all patients of a hospital, for which the same test procedure was executed. This allows to set limits on the values to be expected from

¹ Corresponding Author, Murat Sariyar, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: murat.sariyar@bfh.ch.

a test, for example in the form “the cardiac troponin values on our devices for 99% of all our patients in the last year range between 0.04 ng/L and 100 ng/L”. This type of validation is often called extreme value analysis. For longitudinal assessments, several measurements of the same parameter for the same patient are compared with each other. Plausibility in this context defines whether delta changes are within sensible limits or not. The phrase used for such assessments is “delta check”.

Our working environment was the Institute of Clinical Chemistry (ICC) of Zurich University Hospital (USZ). Several tools were in use to calculate the percentiles and produce graphical outputs, mainly Excel based spreadsheets with added functionality. The workplace of the scientific staff responsible for the validation was separated from the laboratory premises. This physical distance creates a clear separation between the day-to-day business (laboratory examinations) and the validation process, allowing high flexibility in the adaptations of workflows. In this context, we were approached to tackle the question of whether it is possible to increase the efficiency of the medical validation at ICC by business intelligence tools.

2. Methods

2.1 Review of the literature

Our literature research was based on the PubMed and Cochrane Library databases. The focus was on studies that use business intelligence and/or the statistical environment R for delta check procedures and extreme value analysis. Keywords were identified in particular through preliminary discussions with stakeholders. These were then entered into the database mask in various combinations:

- ("medical validation") AND "r statistics"
- ("medical validation") AND "business intelligence"
- ("delta check") OR "r statistics"
- ("delta check") OR "business intelligence"
- ("extreme value analysis") AND "r statistics"
- ("extreme value analysis ") AND "business intelligence"

2.2 Approaches and tools considered

For deriving requirements, we used user stories as well as uses case diagrams. The former technique stems from the agile development framework Scrum [5]. A user story informally describes features of a software system from the perspective of the user; hence, it focuses on what the user needs instead of what the system should deliver. One example in our context: User A (who) wants to calculate percentiles (what) to decide whether validation is necessary (goal). This made it possible to gain a better mutual understanding of the system [6]. Use Cases in contrast describe how the system will act. The creation of a use case was intended to concretize the functional and non-functional requirements of the system and to illustrate the interaction between user and system.

Based on the results of our requirement analysis, we developed a business intelligence application (BIA) using on Microsoft Power BI (Desktop version August 2018) and R (3.5.2). For user-system interactions and visualization, we used an

interactive HTML widget in order to improve the user experience. To implement this widget, additional R packages are required, namely *ggplot2* and *plotly*. The availability of Microsoft Power BI at the ICC and its connection to hospital wide data warehouse made it a natural choice as BI environment.

3. Results

The results of our literature research demonstrated, albeit various methodological improvements were proposed (e.g. see [7]), that there is a gap between the availability of business intelligence tools and their usage for supporting the medical validation workflow in laboratories. In addition, information on the use of expert systems within the clinical laboratory [8], on automated validation for clinical laboratory test results [9], and on the usage of Microsoft Power BI and R for business intelligence applications could be collected [10, 11].

Together with our stakeholders, we defined different workflows to be supported in shape of user stories:

- “As a research assistant, I need up-to-date clinical laboratory test data in order to perform medical validation.”
- “As a research associate, I want graphical visualizations to give me a first idea of the data shape.”
- “As a research associate, I need to calculate percentiles in order to decide whether validation is necessary.”
- “As a research assistant, I need an intuitive user interface to get around quickly.”
- “As a research assistant, I need a user manual so that I can inform myself about the individual functions of the application.”

Discussing these user stories and use case diagrams developed in the requirement analysis phase allowed us to identify following functional requirements:

- i. Interfacing to stored clinical data; in our case to clinical data in the data warehouse of the hospital,
- ii. Support for two separate validation workflows: extreme value analysis and delta check,
- iii. Possibility to specify any time span,
- iv. Filter functions to specify age ranges and sex of the population,
- v. Summary statistics and interactive visualizations (e.g., histograms)
- vi. Automated computation of several percentiles (.01, .025, .05, .95, .975, .99)

Implementing these requirements with a business intelligence software environment allowed us to digitize steps in the validation process that were previously done manually in Excel worksheets; especially, importing the data, calculating percentiles, and producing graphical outputs. In first validations of the tool, the average time for each validation was around two minutes in contrast to 30 minutes with the old solution. Due to the flexibility of the validation workflow, further adaptations can be realized rather quickly. The user interface for the extreme analysis uses filters for

selecting the population on the left panel of the page. The user can specify the laboratory test, the time period, sex and the age range. On the main panel, the results are interactively updated according to the specifications; the core result elements are an interactive histogram, the number of examinations considered, and percentiles. The user interface for the delta check was structured in the same way (see Figure 1). In addition, we included buttons for absolute (default) and relative values (the difference divided by the first value).



Figure 1: The Delta Check page of our application with filters on the left panel and results on the main panel.

Discussion

We were able to increase the efficiency of the medical validation process at the ICC by automating many time-consuming validation steps that were performed manually before. Due to the automation, reduced error rates can be expected. One important advantage of using an R-based BI tool is related to the out-of-the-box functionalities it provides, e.g., for data import, for interactive visualizations and for statistical summaries. The functional requirements list in the result section describes key requirements that typically arise in such contexts.

Development took place on a development environment. Before moving to the production environment, a few hurdles have to be considered. First, there are different Power BI versions: we used Power BI Service for development, but for security reasons (if cloud solutions are not permitted) Power BI Reporting Services should be used, which might require license update. Second, HTML widgets in one Power BI version might not work in another. Third, the graphics integrated into Power BI via R can only deal with data frames having at maximum 150,000 rows [12].

Using R comes with a lot of functionality, stemming from an open research community [13,14], that can be used for advanced statistics and machine learning (ML). Although Python gains high attention for data science applications, for visualization

and scientific-based data analysis, R is still one of the most widespread and dynamic programming environments. The use of ML in a future version could lead to an even faster validation process. The whole process of plausibility checks based on different percentiles and patient characteristics can be automated and trained by a neural network, which can be enabled to integrate many different kinds of information. The keras-package in R [15] is highly suitable for that purpose and its integration will be our next goal.

References

- [1] Hallbach J et al. Klinische Chemie und Hämatologie für den Einstieg. In: Klinische Chemische Hämatologie für den Einstieg. Thieme Stuttgart. 2006;380–383.
- [2] Jennings J et al. Recommended principles and practices for validating clinical molecular pathology tests. Arch Pathol Lab Med. 2009;133(5):743-55.
- [3] Hallbach J et al. Klinische Chemie und Hämatologie. In: Klinische Chemische Hämatologie - Biomedizinische Analyse Für MTLA Studierende. Thieme, Stuttgart. 2011;74–375.
- [4] Nosanchuk JS, Gottmann AW. CUMS and delta checks. A systematic approach to quality control. Am J Clin Pathol. 1974;62:707–712.
- [5] Cervone HF. Understanding agile project management methods using scrum. OCLC Syst Serv. 2011;27(1):18-22.
- [6] Ralph P. The Sensemaking-Coevolution-Implementation Theory of software design. Sci Comput Program 2015;101:21–41.
- [7] Park SH et al. New Decision Criteria for Selecting Delta Check Methods Based on the Ratio of the Delta Difference to the Width of the Reference Range Can Be Generally Applicable for Each Clinical Chemistry Test Item. Ann Lab Med. 2012;32:345-354.
- [8] Valdiguié PM. Role and use of expert systems within the clinical laboratory. Clin Chim Acta. 1998;278:193–202.
- [9] Oosterhuis WP, Ulenkate HJ, Goldschmidt HM. Evaluation of LabRespond, a New Automated Validation System for Clinical Laboratory Test Results. Clin Chem. 2000;46:1811-1818.
- [10] Aspin A. Pro Power BI Desktop. 2nd ed. Apress New York 2017.
- [11] Clark S. Beginning Power BI. 2nd ed. Apress New York 2017.
- [12] Aspin A. High Impact Data Visualization with Power View, Power Map and Power BI. Apress New York 2014.
- [13] Nosek BA et al. Promoting an open research culture. Science. 2015;348:422–1425.
- [14] Tippmann S. Programming tools: adventures with R. Nature. 2015;517:109– 110.
- [15] Ye JJ. Construction and Utilization of a Neural Network Model to Predict Current Procedural Terminology Codes from Pathology Report Texts. J Pathol Inform. 2019;19(eCollection):13.

Using Unsupervised Learning to Identify Clinical Subtypes of Alzheimer's Disease in Electronic Health Records

Nonie ALEXANDER^{ab1}, Daniel C. ALEXANDER^c, Frederik BARKHOF^{cdefg}, and Spiros DENAXAS^{abh}

*a*Institute of Health Informatics, University College London, London, UK,
*b*Health Data Research UK London, *c*Centre for Medical Image Computing,
Department of Medical Physics and Biomedical Engineering, University College
London, London, UK, *d*UCL Institute of Neurology, University College London,
London, *e*UK/Translational Imaging Group, UK/Department of Radiology and
Nuclear Medicine, *f*VU University Medical Center, Amsterdam, The Netherlands,
*g*National Institute for Health Research University College London Hospitals
Biomedical Research Centre, London, UK, *h*Alan Turing Institute, London UK,

Abstract. Identifying subtypes of Alzheimer's Disease (AD) can lead towards the creation of personalized interventions and potentially improve outcomes. In this study, we use UK primary care electronic health records (EHR) from the CALIBER resource to identify and characterize clinically-meaningful clusters patients using unsupervised learning approaches of MCA and K-means. We discovered and characterized five clusters with different profiles (mental health, non-typical AD, typical AD, CVD and men with cancer). The mental health cluster had faster rate of progression than all the other clusters making it a target for future research and intervention. Our results demonstrate that unsupervised learning approaches can be utilized on EHR to identify subtypes of heterogeneous conditions.

Keywords. Phenotyping, Alzheimer's disease, Electronic health records, machine learning

1. Introduction

Alzheimer's disease (AD) is a highly heterogeneous disease. Any two individuals with the disease display a different array of symptoms or progression rate[1]. Progression rate can be increased by many factors such as higher education level and comorbidities like diabetes[2]. Gaining a complete view of a patient's profile of symptoms, comorbidities and demographic factors can enable the discovery of different progression patterns and to personalise treatments for AD patients.

Almost all subtyping studies to date in AD focuses on two data types; cognitive tests[3] and brain scans[4], showing cluster(s) of hippocampal atrophy and more diffuse atrophy, corresponding to specific memory problems and more global cognitive

¹ Corresponding Author, Nonie Alexander, Institute of Health Informatics, University College London, London, UK; E-mail: nonie.alexander.16@ucl.ac.uk

problems, leading to different patterns of progression[4]. Subtyping using electronic health records (EHR) has been carried out in many diseases[5][6] and are ideal data sets to apply these methods to due to the large sample size and the breadth of clinical information[7] they contain. Variables and outcomes are lifted directly from clinical data making them directly relevant to the patient's clinical management. This work aims to identify subtypes of AD using a range of clinical information on the patients.

2. Methods

2.1. Data and cohort

We used data from the Clinical Practice Research Datalink (CPRD)[8] which contains longitudinal primary care EHR from general practices in the United Kingdom[8]. Only patients from practices which have been marked up to standard are used and only data collected after the practice was found to be up to standard were used. Data were extracted and phenotypes defined using the CALIBER data resource[9,10]. We classified patients as cases if they had at least one AD Read code and no other future diagnosis indicating a different dementia subtype diagnosis[11]. Full case analysis was used. This study was approved to use CPRD data by the Independent Scientific Advisory Committee (ref. 18_111).

2.2. Variables and Outcomes

Three categories of variables were included in the analysis; symptoms, comorbidities and demographic and lifestyle factors. Symptoms and comorbidities were identified through a systematic literature review of studies identifying symptoms or finding associated comorbidities. There resulting symptoms found were memory problems, confusion, neuropsychiatric problems and motor problems. From them the disease that were identified were atrial fibrillation, anxiety, hyperglycaemia, rheumatoid arthritis, stroke, hearing loss, depression, kidney disease, heart failure, atherosclerosis and cancer. Age of onset, gender, drinking status and smoking status were all included as demographic variables in the analysis. The phenotypes for each symptom was ideally defined using previously defined and verified CALIBER phenotypes[9,10], if they were not available, using a definition from a previous studies.

We defined five clinical outcomes which were used to evaluate clusters: a) length of time to treatment discontinuation[12], b) rate of progression measured by the Mini Mental State Exam[13], c) healthcare utilization defined as number of appointments and missed appointments per year, d) all-cause mortality and time to assisted living, and e) time to assisted living. For the latter two, we created Kaplan-Meier survival curves.

2.3. Statistical analysis

We used Multiple Correspondence Analysis (MCA) for dimensionality reduction[14]. K-means was used to identify clusters: First, k was decided through an elbow plots for the cluster entropy, silhouette coefficient and Bayesian Information Criterion (BIC). K-means was executed 100 times to establish the optimum solution determined by the

inter cluster variation. To evaluate the clusters silhouette score were used to measure cluster structure and cluster stability was measured through Jaccard index of cluster results based on a 100 X bootstrapped sample.

3. Results

We identified 7,913 AD patients (66.2% female, mean age 82.1, 81.3-83 95% CI). Using the first five MCA principal components, we identified five clusters: mental health cluster which has the highest prevalence of anxiety and depression comorbidities, non-typical AD which had the lowest prevalence of memory symptoms and highest of other symptoms, typical AD which had high prevalence of memory symptoms and low other symptoms, cardiovascular disease (CVD) cluster with high prevalence of cardiovascular diseases such as atrial fibrillation and associated diseases such as diabetes and a cluster of men with cancer (Table 1, Figure1) all with differing clinical outcomes (Figure 2). The clusters had a silhouette score of 0.19 (showing weak cluster structure) and mean Jaccard index of 0.78 indicating overall stability.

Table 1. Each cluster name, key variables that characterise the cluster and their outcomes. Green - highest outcome. red - lowest outcome, * significant

Cluster I.D	Demographics	Top Characteristics (figure 1)	Outcomes (figure 2)
1. Mental Health Cluster	n = 1528 80.1% female onset age = 75 Diag MMSE = 21.6	Neuropsychiatric symptoms, (83.51%) Anxiety (62.11%) Depression (69.24%)	Fastest rate of progression* (2.16 mmse points per year, 2.06-2.26 95% CI)
2. Non- typical AD Cluster	n = 1640 81.3% female onset age = 86 Diag MMSE = 20.9	Memory symptoms (44.15%) Confusion(71.04%) Neuropsychiatric symptoms (98.35%)	Shortest time till treatments discontinuation (3.09 years, 3.19-2.99 95% CI) Lowest survival probability* Shortest time till assisted living *
3. Typical AD Cluster	n = 2026 90.1% female onset age = 82 Diag MMSE = 21.7	Memory symptoms (89.09%) Confusion(0.05%) Neuropsychiatric symptoms (18.85%)	Fewest appointments per year* (16.6 apps per year, 16.15-17.1, 95% CI)
4. Cardiovascular disease Cluster	n = 686 45.9% female onset age = 81 Diag MMSE = 21.4	Diabetes (84.26%) Hypertension(81.49%) Atrial fibrillation(20.41 %)	Greatest number of appointments per year* (24.8 apps , 23.7-25.9 95% CI) and missed appointments per year* (0.96 missed apps, 0.84 - 1.08 95% CI)
5. Memory problems and Cancer Cluster	n = 1710 19.0% female onset age = 79 Diag MMSE = 22.6	Memory symptoms (90.94%) Cancer(43.27%) Male (80.99%)	Slowest progression rate (1.59 mmse points a year, 1.5-1.68 95% CI)

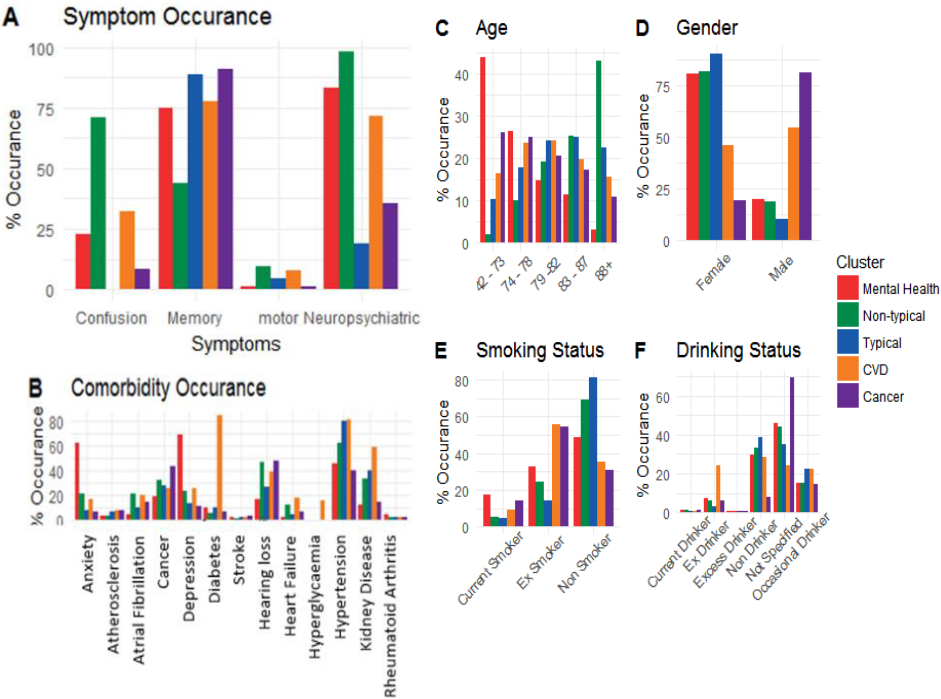


Figure 1: Distribution of the variables included in the cluster analysis for each cluster a) symptoms, b) Comorbidities, c) age, d) gender, e) clinically-recorded smoking status, f) clinically-recorded drinking status

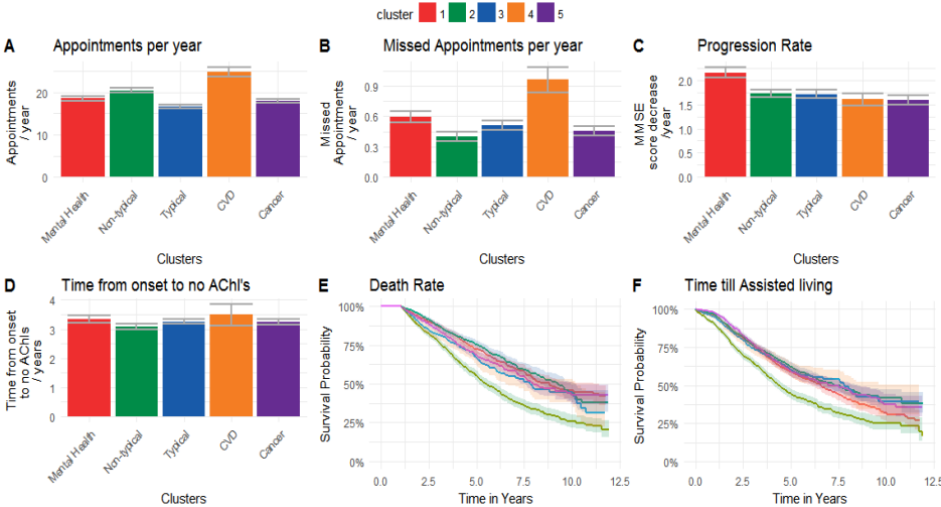


Figure 2 : Outcomes difference for variables not included in the cluster analysis per cluster: a) mean number of appointments per year with CI, b) mean missed appointments per year with 95% CI, c) mean MMSE score reduction per year with 95% CI, d) mean time to treatment discontinuation with 95% CI, e) Kaplan Meyer curve of survival probability, f) Kaplan Meyer curve on time till assisted living.

4. Discussion

Using unsupervised machine learning on a cohort of 7,913 AD patients defined through EHR, we identified five distinct clusters with different clinical profiles: a mental health cluster, a non-typical AD cluster, a typical AD cluster and clusters with mostly cardiovascular problems and a cluster with cancer. The mental health and early onset cluster had a faster rate of progression. Further research is needed to delineate the relationship between the mental health comorbidities, rate of progression and early onset. Identifying the CVD cluster can have clinical benefit as they have the highest healthcare utilization. The two clusters of typical and non-typical AD are similar to clusters found in previous research splitting patients with memory issues and patients with other cognitive issues[3]. The cluster with high prevalence of cancer, and men reported low prevalence of other symptoms aside from memory loss. There should be more investigation into whether the mechanism that leads cancer to protect against AD [2] also reduces the number of symptoms through less global atrophy. Using this data type uses a unique perspective on which to cluster patients as it offers a longitudinal and broad scope of a patient. The cluster results also have clinically relevant and impactful outcomes. There are however several weaknesses to the study such as there are some that are important but not or only partly recorded such as family history of AD recorded in EHR. Future work will look into validating these results through comparing this method with other cluster methods cluster results using imaging and cognitive test results.

References

- [1] D. Mungas et.al, The effects of age on rate of progression of Alzheimer disease and dementia with associated cerebrovascular disease, *Arch. Neurol.* **58** (2001) 1243–1247.
- [2] J.-Q. Li, et.al, Risk factors for predicting progression from mild cognitive impairment to Alzheimer's disease *Journal of Neurology, Neurosurgery & Psychiatry.* **87** (2016) 476–484. d
- [3] N.M.E. Scheltens, et.al, Cognitive subtypes of probable Alzheimer's disease robustly identified in four cohorts, *Alzheimers. Dement.* **13** (2017) 1226–1236.
- [4] A.L. Young, et.al Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference, *Nat. Commun.* **9** (2018) 4273.
- [5] L. Li, et.al, Identification of type 2 diabetes subgroups through topological analysis of patient similarity, *Sci. Transl. Med.* **7** (2015) 311ra174.
- [6] T. Ahmad, et.al, Clinical implications of chronic heart failure phenotypes defined by cluster analysis, *J. Am. Coll. Cardiol.* **64** (2014) 1765–1774.
- [7] H. Hemingway et.al, Big data from electronic health records for early and late translational cardiovascular research: challenges and potential, *Eur. Heart J.* **39** (2018) 1481–1495.
- [8] E. Herrett, A.M. Gallagher, K. Bhaskaran, H. Forbes, R. Mathur, T. van Staa, and L. Smeeth, Data Resource Profile: Clinical Practice Research Datalink (CPRD), *Int. J. Epidemiol.* **44** (2015) 827–836.
- [9] S. Denaxas, et.al, UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER, *J. Am. Med. Inform. Assoc.* (2019). doi:10.1093/jamia/ocz105.
- [10] V. Kuan, S. Denaxas, A. Gonzalez-Izquierdo, K. Direk, O. Bhatti, S. Husain, S. Sutaria, M. Hingorani, D. Nitsch, C.A. Parisinos, R.T. Lumbers, R. Mathur, R. Sofat, J.P. Casas, I.C.K. Wong, H. Hemingway, and A.D. Hingorani, A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service, *Lancet Digit Health.* **1** (2019) e63–e77.
- [11] M. Pujades-Rodriguez, et.al, The diagnosis, burden and prognosis of dementia: A record-linkage cohort study in England, *PLoS One.* **13** (2018) e0199026.
- [12] National Institute for Health and Care Excellence (UK), Dementia: Assessment, management and support for people living with dementia and their carers, London, 2018.
- [13] S. Dubin, The Mini Mental State Exam, *The American Journal of Nursing.* **98** (1998) 16D.
- [14] B. Le Roux, and H. Rouanet, Multiple Correspondence Analysis, SAGE, 2010.

Assessment of the Prognostic Accuracy of Scores in Different Ways on the Example of Assessing the Validity of the Prognostic Model ISS-RTS-TRISS

Irina V. VASILYEVA^{a,1b}, Tatyana V. ZARUBINA^b, Yuri D. UDALOV^a

^a State Research Center - Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russia

^b The Russian State Medical University, Moscow, Russia

Abstract. The aim of the present study was a comparison of prognostic accuracy assessment results for scores in different ways on the example of validity analysis of the prognostic model ISS-RTS-TRISS for assessing the severity of the condition in children with trauma. The prospective study was conducted using clinical and physiological data collected at the admission and during the first 24 hours of hospitalization from 414 children with trauma. We had three groups of patients, common group children with traumatic brain injuries, and two groups of patients were divided into the two following ways: 141 (34%) patients with isolated traumatic brain injuries and 273 (66%) patients with combined injuries with traumatic brain injuries. The validity and prognostic accuracy of prognostic scores were assessed by determining their discrimination and calibration ability. Analysis of the discrimination ability of score was carried out by assessment the areas under the ROC curves. For analysis of the calibration ability of score was used three methods of the Hosmer-Lemeshow test (H- criterion, C-criterion in two variants). The ISS-RTS-TRISS score showed significantly outstanding predictive accuracy in studied groups (AUROC ≥ 0.9). However, estimation the calibration ability of score using the C-criterion, by dividing into groups with the same number of cases of patients (recommended for abnormal distribution) did not show an unambiguous results. It was shown that to obtain an unambiguous correct result, it was necessary to use the C-criterion method using the division of cases into groups with the same number of patients with the lethal outcome. When using this method, satisfactory results of study of calibration ability were shown for all the studied groups ($p < 0.05$).

Keywords. Prognostic score, validity of prognostic scores, prognostic accuracy of score, the Hosmer-Lemeshow test, score ISS-RTS-TRISS

1. Introduction

Modern medicine focuses on increasing the effectiveness of decisions taken during treatment, based on assessing the severity of the condition and the severity of injury in patients, predicting the likelihood of death in patients [1, 2]. One of the methods for

¹ Irina V. Vasilyeva, State Research Center - Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Marshala Novikova Str. 23, Moscow, 123098, Russia, iv001yt@gmail.com

solving this problem is the using of numerous specialized prognostic scores for various nosologies and different age groups of patients [3, 6, 7, 8]. The prognostic scores, which are mathematical models that include basic demographic, clinical and laboratory data, are an example of the quantitative assessment of the severity state [8, 9, 10]. One of these scores is the ISS-RTS-TRISS (Injury Severity Score - Revised Trauma Score - Trauma Injury Severity Score) prognostic score for assessing the severity of injuries, proposed by Champion and colleagues in 1990. The score allows to assess the severity of a patient's injury and determine the likelihood of death [11]. An important point in the use of scores is the quality and validity of the predictive model, which determines the validity and suitability of the application of the technique in specific conditions, and consequently - the correspondence of the possibility of prognostic scores to the tasks assigned [5, 12, 13]

As is known, the validity and prognostic accuracy of prognostic scores can be assessed by determining its discrimination and calibration ability [3, 4, 5]. Determination discrimination ability at the moment is not difficult in the presence of matching the amount of data.

However, the determination of the calibration ability is a more complicated process and depends on the approach chosen for calculating the Hosmer–Lemeshow test [4, 5].

The aim of the present work was a compare of prognostic accuracy assessment results for scores in different ways on the example of analysis the validity of the prognostic model ISS-RTS-TRISS for assessing the severity of the condition in children with traumatic brain injuries (TBI), located in the department of anesthesiology and resuscitation in the Clinical and Research Institute of Urgent Pediatric Surgery and Trauma. To achieve this aim we set following tasks to: investigate discrimination ability of the ISS-RTS-TRISS prognostic score in groups with TBI and investigate the calibration ability of the ISS-RTS-TRISS by three ways of Hosmer–Lemeshow test.

2. Material and methods

The prospective study was performed on basis of data collected from 414 patients (mean age 9.3 ± 5 , m/f were 266/148 patients) with TBI from the age of 1 month to 18 years old who had been admitted to the Department of Anesthesiology and Resuscitation between January 2008 and January 2012 of Clinical and Research Institute of Urgent Pediatric Surgery and Trauma. 38 (9.2 %) patients died.

Criteria for including patients into the study were the following: traumatic brain injury, age till 18, availability of all necessary data for calculating results of prognostic scores, staying in the anesthesiology and resuscitation department not less than 24 hours.

We had three groups of patients, common group children with traumatic brain injuries, and two groups of patients were divided into the two following ways: 141 (34%) patients with isolated traumatic brain injuries and 273 (66%) patients with combined injuries with traumatic brain injuries (Table 1).

The assessment of severity of patient's condition and injury severity and the death risk were determined with ISS-RTS-TRISS prognostic score; these data were collected from anamnesis, initial examination, parameters of physiological and neurological

status on admission and within the first 24 hours. Score was calculated using software tools published on the Société Française de Réanimation community website [14].

Table 1. Characteristics of data for all patients and in different groups

Group patients with	N	Age (m± δ)	M/F (%/%)	Died (%)	Survived (%)
TBI	414	9.3±5	266/148 (64.3/35.7)	38 (9.2)	376 (90.8)
Isolated TBI	141	9.0±4.9	88/53 (62.4/37.6)	10 (7.1)	131 (92.8)
Combined injuries with TBI	273	9.1±4.8	171/102 (62.6/37.4)	25 (9.2)	248 (90.7)

N - sample's size of group, m - average age in the group, δ - standard deviation of mean, M – man-child, F – female-child, Died – the number of deaths in the group of patients, Survived - number of surviving patients in the group, TBI - traumatic brain injuries.

The validity and prognostic accuracy of prognostic score were assessed by determining its discrimination and calibration ability. The discrimination ability has shown a probability score to divide patients into two groups: with favorable outcomes and unfavorable (likelihood and lethal) outcomes. ROC-curve was used for evaluating the discrimination ability of the studied prognostic score [4, 5]. For analysis of the calibration ability of score was used three methods of the Hosmer-Lemeshow test. The first way was the H-criterion test with the division patients 10 groups by probability of lethal outcome. In this case, the number of patients in the intervals according to the probability of a lethal outcome wasn't the same; the number of patients who died in the intervals was also being different. The second way was the classic C-criterion of Hosmer-Lemeshow test with the division patients into probability groups of death with the same number of observed cases of patients. And the third way – the C-criterion of Hosmer-Lemeshow test with the division patients into probability groups of death with the same number of observed cases with lethal outcome of patients [5]. The data base statistical assessment and statistical description were done in Excel of Microsoft Office 2013 and IBM SPSS Statistics 20 programs.

3. Results

The investigated prognostic score ISS-RTS-TRISS showed significantly outstanding discrimination ability in studied groups (AUROC>0.9). [14]. (Figure 1, Table 2).

Table 1. Results of AUROC and cut-off of prognostic score Characteristics of data for all patients and in different groups

Group patients with	N	AUROC	p
TBI	414	0.928±0.019	<0.001
Isolated TBI	141	0.985±0.012	
Combined injuries with TBI	273	0.899±0.029	

N - number of patient in the group, AUROC –area under ROC curve, δ- standard deviation, TBI - traumatic brain injuries.

Calibration ability for score ISS-RTS-TRISS determined by H-criterion of the Hosmer-Lemeshow test showed satisfactory calibration ability for all considered groups (p<0,05) (Table 3). The classic C-criterion of Hosmer-Lemeshow test (division group into the same number of cases of patients) didn't show satisfactory calibration ability in all considered groups (p>0,05). The second C-criterion of Hosmer-Lemeshow test (division group into the same number of patients with a lethal outcome) showed satisfactory calibration ability for all groups (p<0,05). The second C-criterion of Hosmer-Lemeshow method proved to be insufficiently sensitive to assess the calibration ability of the scores with small number cases of study. Thus, the observance

of the conditions for the need for a sufficient amount of not only the general group, but also the need for a sufficient amount of the studied subgroups, taking into account favorable and unfavorable outcomes.

In general, it can be concluded satisfactory calibrating ability for all considered groups of patients with TBI.

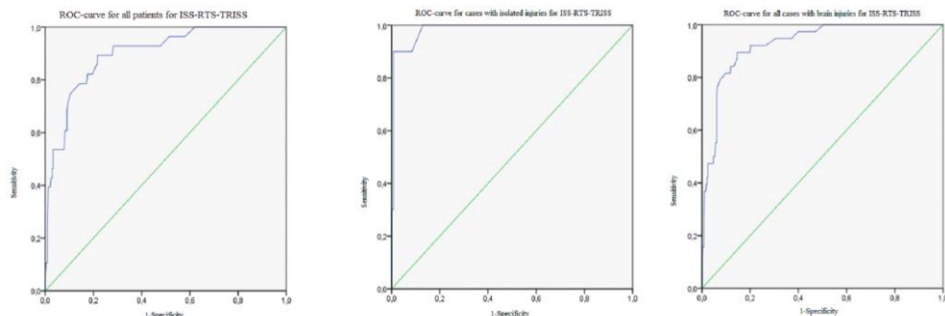


Figure 1. ROC-curves for prognostic score ISS-RTS-TRISS in studied groups.

Table 3. Results of assessment of calibration ability for prognostic score ISS-RTS-TRISS in studied groups.

Studied groups of patients with	N	H-criterion			C-criterion (the same number of cases of patients)			C-criterion (the same number of patients with a lethal outcome)		
		χ^2	p	f	χ^2	p	f	χ^2	p	f
TBI	414	26.04	0.001	10	14.91	0.06	10	47.75	<0.001	10
Isolated TBI	141	20.26	0.009	10	3.18	0.786	8	8.71	0.033	5
Combined injuries with TBI	273	26.68	<0.001	10	15.20	0.055	10	24.09	0.002	10

N- number of patient in the study, χ^2 - χ^2 -value, p-p-value (Hosmer–Lemeshow test), f-degree of freedom, TBI- traumatic brain injuries.

This analysis of the prognostic accuracy of scales in different ways by using three methods of the Hosmer-Lemeshow test on the example of assessing the validity of the prognostic model ISS-RTS-TRISS has shown, that for sufficient validity of the results there must be, not only the total number of survived patients, but also must be a sufficient number of dead patients in the study group.

4. Discussion

Modern scientific evidence-based medicine in its practice, more and more uses complex mathematical statistical methods of analyzes.

Thus, for investigation of prognostic accuracy of prognostic scores used the determination of discrimination and calibration abilities of mathematical models. The aim of the present work was to compare of results of assessment and prognostic accuracy of scales in different ways on the example of analysis the validity of the prognostic model ISS-RTS-TRISS for assessing the severity of the in children with traumatic brain injuries.

For validated prognostic score ISS-RTS-TRISS were defined discrimination ability (AUROC >0,9) and calibration ability by Hosmer-Lemeshow test in three ways in considered groups.

Our investigation has shown satisfactory calibrating ability for all considered groups by H-criterion of Hosmer-Lemeshow test ($p < 0,05$).

The classic C-criterion of Hosmer-Lemeshow test with the division patients into probability groups of death with the same number of observed cases hasn't shown satisfactory calibration ability in all considered groups ($p > 0,05$). The C-criterion of Hosmer-Lemeshow test with the division patients into probability groups of death with the same number of observed cases with lethal outcome has shown satisfactory calibration ability for all groups ($p < 0,05$). In general we could conclude calibrating ability for all considered groups was satisfactory.

Thus, as a result of our study, it was shown that for assessment of calibration ability the Hosmer-Lemeshow test should be used, but one should take into account the number of dead and survived patients, as well as their distribution in the interval groups.

References

- [1] A. Cook, J. Weddle, S. Baker, D. Hosmer, L. Glance, L. Friedman, T. Osler, A comparison of the Injury Severity Score and the Trauma Mortality Prediction Model. *J TraumaAcute Care Surg* 2014; 76(1): 47–53
- [2] M. Letts, D. Davidson, P. Lapner, Multiple trauma in children: predicting outcome and long-term results, *Can. J. Surg.* – 2002. – Apr;45(2). –P.126. – PMID: 11939656.
- [3] I.V. Vasilyeva, S.L. Shvirev, T.V. Zarubina, O.V. Karaseva, A.S. Samoylov, YU.D. Udalov, [Estimation of the Accuracy of Prognostic Scores for the Treatment of Children with Severe Trauma in a Specialized Trauma Hospital](#). *Stud Health Technol Inform* 2018 ;248:263-269
- [4] S. Lemeshow, D.W. Hosmer, A review of goodness of fit statistics for use in the development of logistic regression models, *Journal Epidemiology* 115 (1982), 92- 106
- [5] D. W. Hosmer& S. Lemeshow - *Applied Logistic Regression* - 2nd Edition 2000, New York:Wiley.
- [6] M. Letts, D. Davidson, P. Lapner, Multiple trauma in children: predicting outcome and long-term results, *Can. J. Surg.* – 2002. – Apr;45(2). –P.126. – PMID: 11939656.
- [7] P. Bhadoria, A.G. Bhagwat, Severity Scoring Systems in Paediatric Intensive Care Units, *Indian Journal of Anaesthesia.*- 2008.-52: Suppl (5).- P.663.
- [8] H. R.Champion, Trauma scoring, *Scandinavian Journal of Surgery.*- 2002. – 91(1). – P.12.
- [9] C.K. Senkowski, M.G. McKenney, Trauma scoring systems: a review, *J American College of Surgeons*, 1999, Nov; 189(5): 491-503, Qureshi,
- [10] S.P. Baker et al. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care, *Journal Trauma*14 (1974), 187-96.
- [11] C.R. Boyd et al., Evaluating Trauma Care: The TRISS Method, *Journal Trauma* 27 (1987), 370-8.
- [12] W.A. Knaus, E.A. Drapper, D.P.Wagner et al. APACHE II: a severity of disease classification system, *Journal Critical Care Medicine* 13 № 10 (1985), P. 818–829.
- [13] J.J. Tepas, D.L. Molitt, J.L. Taibert and M.Bryant, The Pediatric Trauma score as a predictor of injury severity in the injured child, *Journal Pediatric Surgery* 22 (1987), 14-18
- [14] <https://sfar.org/espace-professionnel-anesthésiste-reanimateur/outils-professionnels/scores-sfar/> (Access during 2008-2016 years, last access 10.10.2019).

VIVID: Independent Living of Aging Adults Suffered a Stroke

Athanasios ANASTASIOU^a, Ourania MANTA^a, Dimitra ILIOPOULOU^a and
Dimitrios KOUTSOURIS^a

^aBiomedical Engineering Laboratory, National Technical University of Athens,
Athens, Greece

Abstract. Based on the recent statistics published by the Stroke Association (UK), first-time incidence of stroke occurs almost 17 million times a year worldwide (one every two seconds), making Stroke as the second cause of death in the world. By the age of 75, 1 in 5 women and 1 in 6 men will have a stroke, which is one of the largest causes of disability, as half of all stroke survivors have a disability, making those persons dependent on others (1 in 5 are cared for by family and/or friends). People living longer is a cause for celebration, but older people are more vulnerable to mental health, cognition and physical problems, especially if they have already experienced a stroke (minor or mild). Depression is a main condition after a stroke and may be experienced in the form of sadness, unexplained pains, loss of interest in socializing, weight loss etc. The abovementioned conditions reduce the person's ability to remain active and independent, affecting their well-being and quality of living. Independent living of aging adults that have suffered a stroke is the key motivation for the VIVID project.

Keywords. stroke; machine learning; personalization; virtual coach ; serious gaming; affective computing

1. Introduction

Over the past decade and based on the undeniable fact that Europe is an ageing continent, the ICT research community has responded with enthusiasm to the latest challenge: providing Europeans with healthier, safer, more independent and overall happier ageing. As Baby Boomers age, the need for adult care and assisted living continues to grow. As of February 2015, 16.4 million people in Europe receive support annually from five main long-term care service options: home health agencies, nursing homes, hospices, residential care communities, and adult day service centres. Outside these facilities, 43.5 million family and informal caregivers currently provide care for someone aged 50+.

Among the most severely disabled old persons living in Europe, about two-thirds rely solely on family members for assisted care, often resulting in great emotional stress for family caregivers. The aging population has greater chances to experience age related conditions, such as Stroke and the Post Stroke Depression. We aim to utilize intelligent multisensory and multimodal ICT devices to sense, monitor and understand human activity, in order to motivate the aging population to have a better quality of living, while improving their physical, cognitive, mental and social well-being, and furthermore to support them in their everyday living tasks. We are proposing a universal solution for any aged person in the form of a Personalized Virtual Coach (PVC) that will sense mood, activity, physical status and stress of the user (using smart environment sensors, wearable devices and voice analysis), and will interact with the user via intuitive Graphic User Interfaces and Voice Recognition. The PVC will reside in the Home Hub of the VIVID

platform, while it will be accessible via smartphone and/or tablet devices. Our solution will adopt the latest development in Human Machine Interaction and Affective Computing, to provide user-adapted and intuitive access to a multitude of enabling computer applications, designed to assist and support the welfare of the seniors. The PVC aims to reduce the decline of the physical status, the mental capacity and the cognitive abilities of seniors suffered a Stroke, including those having Post Stroke Depression symptoms, encouraging them to stay active and perform physiotherapy and speech therapy rehabilitation sessions, using body posture tracking devices and Augmented Reality headsets. Going beyond any other similar solution, the proposed PVC prototype will incorporate an ontology-based adaptive learning system to improve machine-to-machine learning, in order to enhance the identification of the user behaviour, patterns and preferences, and provide a highly customised virtual coaching experience.

2. Objectives

2.1. Develop the Smart Sensor Network (SSN) to capture the living environment and the health status of the user

Under this objective, we will investigate how we can use Augmented Reality, Serious Games, Voice Analysis and Emotional Computing to monitor how the user interacts with the living environment in their everyday living conditions. The proposed framework will be capable of using Internet of Things (IoT) enabled commercial sensors to a) "sense" the home environment of the user, and b) measure physiological parameters to assess the health status of the user. SSN will adopt UniversAAL middleware to allow the IoT enabled devices to become available, searchable, accessible, and usable by the UniversAAL to enable applications of the system.

2.2. Develop an intuitive Interaction Framework for the interaction between the user and the PVC

This objective will focus on the adoption of the intuitive Human Computer Interaction (HCI) principle in the implementation of the PVC. The interaction will be performed using four of the user's senses (sight, hearing, voice and touch). Portable devices with touch screens will be used to receive user's input, using simple and intuitive Graphic User Interfaces (GUIs). A Natural Language Processing (NLP) framework will be used to process voice commands and reply to user requests. The NLP framework will enable PVC to be built upon the use of conversational interfaces that understand and speak human language, leveraging the acceptance and maximising user friendliness.

2.3. Develop a Serious Gaming Platform supported by Augmented Reality headset

This objective will focus on the development of a Serious Gaming Platform (SGP) to be used by the elderly to perform physical and cognitive exercises. The SGP will be composed of standalone applications installed in the HH, which will update their content and exercises via internet connection. The updated content will be provided by the medical professionals and will be personalised to the individual player, supported by the IF. This module will also be used for the tele-rehabilitation (physiotherapy and speech therapy) of the users.

2.4. Improve prescription adherence along with medication compliance and persistence

Successful treatment of disease with prescription medicines requires consistent use of the medicines as prescribed. Yet research shows that medicines commonly are not used as directed. Non-adherence to medicines is a major health care cost and quality problem, with numerous studies showing high rates of non-adherence directly related to poor clinical outcomes, high health care costs, and lost productivity. Many of the human and economic costs associated with non-adherence can be avoided, making improving patient adherence one of the best opportunities to get better results and greater value from our health care system [7]. Prescription adherence and medication compliance are crucial when targeting Depression and Stroke patients. Forgetfulness and complexity of regimen are the two parameters of adherence where PVC will focus. The proposed Medication Adherence Module (MAM) will be composed of a smart pill box (an IoT connected device to manage medication disposal) and a reminder connected to the user via the IF.

2.5. Improve exercise compliance to improve physical well-being and reduce health related risks

This objective will focus on the personalised provision of the exercise plan for the users, based on the individual preferences and needs. There is clear evidence that well-designed technology interventions can effectively change health-related behaviour, improve patient knowledge and confidence for self-management of health, and lead to better health outcomes. Embedding emerging technologies into health interventions afford a real opportunity to study, understand, and positively affect human behaviour through devices which monitor human behaviour and provide an ideal platform for delivering behavioural change interventions and feedback. By improving the physical health of the seniors via the executed activities, both musculoskeletal and mental well-being will be improved, as strong muscles can reduce the risks of potential falls, and as exercise can treat mild to moderate depression as effectively as antidepressant medication. Validation of the results will be performed using the International Physical Activity Questionnaires (IPAQ).

2.6. Improve cognitive well-being to reduce memory decline, preserve language skills and maintain visual perceptual abilities

Language and vocabulary are well retained throughout the lifespan. Vocabulary continues to improve into middle age. Recall of general knowledge acquired at a young age and well-practiced skills like arithmetic also peak in middle age and are resistant to age-related decline. To be able to recall an event or new information, the brain must register the information, store it, and then retrieve it when needed. The ability to recall new information, such as reading material, becomes more challenging after the age of 40, particularly for visual material. To improve the cognitive skills of aging people, the PVC will challenge the users to play mini game, like quizzes or simple problem solving games, tailored for each individual.

2.7. Maintain mental health of the aging population to enhance the everyday feeling of well-being

Under this objective, the mental health of the seniors will be promoted, as it is an integral part of health. Good mental health is identified by WHO as a "state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and is able to make a contribution to his or her

community." With that in mind, VIVID will try and maintain the improvement of mental health via the guidance and the support of the PVC, the adoption of a healthy diet, the performance of regular moderate exercise , the adoption of patterns that offer good quality of sleep or manage stressful situations (via relaxation techniques). To achieve those targets, the PVC will encourage the user to comply with a minimum set of goals, including regular exercising, setting small incremental targets in everyday life (regarding exercise, sleeping patterns or participation in community activities) to help him/her maintain a competent mental health .

2.8. Advance the social wellness of the seniors, to let them remain active and socially connected

To maintain the social wellness of the seniors, the PVC will encourage their contribution to their environment and community. Social wellness incorporates making healthy living choices and initiating better communication with others. Social activities such as games (bingo, dominos, cards, and puzzles), parties/gatherings for family and friends or the local community, are part of the activities to improve social wellness. By helping seniors to be kept socially active, there is a decreased risk of diseases, social isolation as one of the key indicators leading to depression in seniors is avoided (the seniors maintain and continue to build new relationships). There is also proof that the immune system tends to be stronger in seniors who are active socially, while isolated seniors tend to have higher blood pressure, because the lack of social connections can lead to an elevated risk of diseases, such as cardiovascular disease .

3. Concept and methodology

The concept of the VIVID project is the development of a personalised solution, in the form of a PVC, that will use the latest advances in ICT, to motivate and support the aging

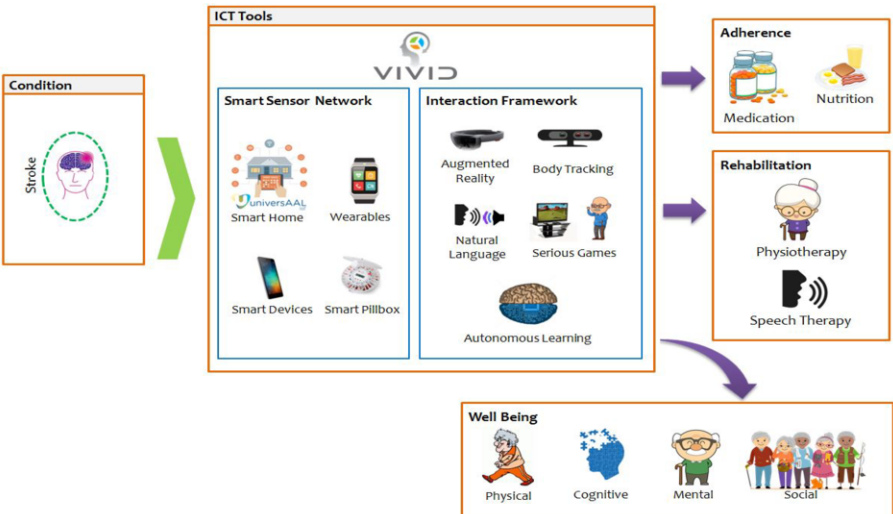


Figure 1 VIVID PVC architecture outline

adults in terms of medication and exercise adherence and their overall well-being. VIVID will develop a personalized solution in the form of a PVC, adapted to the needs of the

aging population, with focus on those with PSD and those who have suffered a SE and need rehabilitation. Our solution targets those that require rehabilitation (speech therapy or physiotherapy) after they have suffered a Stroke Episode. This will be done in such a way that will preserve the individual's dignity and will motivate him/her to re-learn basic skills that the stroke may have affected, such as speaking and walking. Figure 1 depicts the ICT tools that will be used in the proposed PVC solution offered by the VIVID project.

4. Discussions

VIVID can improve care and reduce the need to use health and care services by enabling more of our citizens to build up the knowledge, skills and capabilities they need to manage their own care and symptoms. VIVID will achieve this by engaging communities, supporting carers, and developing volunteering in and partnerships with the charitable and voluntary sector. VIVID will aid in Improvement of patients' life-space mobility (Modified Ashworth scale, Tinetti Performance Oriented Mobility Assessment, High-level Mobility Assessment Tool (HiMAT), in Permission of independent living (e.g. EQ5-D, SF-36, Interprofessional Socialization and Valuing Scale (ISVS), Care Dependency Scale, Bowel Control Scale (BLCS), Paced Auditory Serial Addition Test (PASAT), in Improvement of patients' autonomy and participation in society (Lkert scale), in Improvement of family living standards (self-reported questionnaire), in Improvement of quality of life (questionnaires of QoL – EQ 5D) and in reduction of costs for the management of the gait problems (Functional Gait Assessment). In the VIVID project the users' participation is situated in the center of the development process. Innovative user-friendly interfaces of VIVID will increase the usability and adaptability of the systems. Taking into account seniors' needs, an interactive game framework is developed and will be tested during the pilot phase. The science of behavior and personalization will help in order to build a really elderly-friendly and easily adopted VIVID system.

5. References

- [1] T. S. Association, "State of the Nation," *Stroke Stat*, no. January, pp. 1–40, 2016.
- [2] A. G. Karatepe et al., "Comorbidity in patients after stroke: Impact on functional outcome," *J Rehabil Med*, vol. 40, no. 10, pp. 831–835, 2008.
- [3] M. Liu et al., "Comorbidities in stroke patients as assessed with a newly developed comorbidity scale," *Am J Phys Med Rehabil*, vol. 78, no. 5, pp. 416–24.
- [4] S. Paolucci et al., "Quantification of the risk of poststroke depression: the Italian multicenter observational study DESTRO," *Acta Psychiatr Scand*, vol. 112, no. 4, pp. 272–278, Oct. 2005.
- [5] E. M. Whyte et al., "Post stroke depression: epidemiology, pathophysiology, and biological treatment," *Biol Psychiatry*, vol. 52, no. 3, pp. 253–64, Aug. 2002.
- [6] S. Asteriadis et al., "Towards Detecting Clusters of Players using Visual and Gameplay Behavioral Cues," *Procedia Comput Sci*, vol. 15, pp. 140–147, 2012.
- [7] A. Anastasiou et al., "Monitoring of compliance on an individual treatment through mobile innovations," *XIV Mediterr Conf*, 2016.

This page intentionally left blank

Section 2

Supporting Care Delivery

This page intentionally left blank

A Domain-Independent Semantically Validated Authoring Tool for Formalizing Clinical Practice Guidelines

Jordi TORRES^{a, †}, Garazi ARTOLA^{a, †} and Naiara MURO^{a,b,c}

^a*eHealth and Biomedical Applications, Vicomtech, Donostia-San Sebastian, Spain*

^b*Biodonostia, Donostia-San Sebastian, Spain*

^c*Sorbonne Universités, UPMC Univ Paris 06, INSERM, Université Paris 13, Sorbonne Paris Cité, UMR S 1142, LIMICS, Paris, France*

[†]*Equally Contributing Authors*

Abstract. Clinical Practice Guidelines (CPGs) are promoted as a powerful tool for standardization of the medical care quality and improvement of patients' outcomes. However, CPGs need to be formalized in a computer interpretable format (i.e. as Computer Interpretable Guidelines or CIGs) for their implementation within Clinical Decision Support Systems (CDSS). But, maintaining the reliability of these guidelines when deploying them in different clinical settings is still a challenge. On the one hand, the complexity of the medical language complicates the adoption of the guidelines in different clinical institutions. On the other hand, the continuous discovery of new evidence needs to be included within CPGs, updating their contents and providing tools for evidence assessment. Furthermore, although nowadays' clinical decision-making tends towards a personalized process, guidelines are designed for a general population. In this paper, we present an Authoring Tool (AT) that allows clinicians to take an active role in the process of CPG formalization. This AT enables them to introduce new clinical knowledge and create personalized CIGs for their local application, which best fits their clinical needs. The proposed system also allows the use of ontologies to facilitate the standardization and interoperability of the created guidelines. Finally, the content included in the CIGs can be evaluated using standard systems for grading clinical evidence.

Keywords. Authoring Tool, Clinical Decision Support System, Clinical Practice Guidelines, Computer Interpretable Guidelines, Ontologies.

1. Introduction

Clinical Practice Guidelines (CPGs) are a set of criteria developed in a systematic way to help professionals in the decision-making process, providing the latest evidence-based diagnostic or therapeutic options when dealing with a health problem or a specific clinical condition [1]. Over the past years, CPGs have been widely promoted to be implemented as Computer Interpretable Guidelines (CIGs) within Clinical Decision Support Systems (CDSS). Nevertheless, there is still work to be done in the maintenance and personalization of these guidelines in order to maintain their reliability when implementing them in different clinical settings. On the one hand, the complexity of medical language makes the comprehension and interoperability of the

CPGs a complicated issue, since different guidelines modeling the same clinical domain could differ in the provided knowledge. In this context, semantic web technologies such as ontologies are promoted for a standardized medical vocabulary. On the other hand, the evolving nature of medicine and the continuous discovery of new evidence needs to be included in CPGs, assessing the quality of the formalized evidence in a continuous way. To solve this issue, methodologies that assess the quality of the formalized evidence and the strength of the recommendations are promoted. Furthermore, CPGs are developed considering a population as target, not individuals, thus assuming the existence of a “standard” patient, which is not representative of all possible individual cases [2]. In this sense, personalizing guidelines using clinicians’ own experience, patient preferences or local protocols of institutions can be helpful to increase clinicians’ compliance to CDSS [3].

In this paper, an Authoring Tool (AT) that enables clinicians to formalize actively personalized CPGs in a user-friendly way by introducing new clinical knowledge and create CIGs that can be adapted to their local protocols is proposed. First, an approach for the semantic validation of the content in the guidelines leaning on an ontology is presented. Second, the methodology used for the CPG formalization process is described. Lastly, a use case in Gestational Diabetes Mellitus is shown to illustrate the proposed methodology.

2. State of the Art

The formalization process of CPG into CIGs is a complex, time-consuming task requiring technical and clinical skills to be done successfully. As a result of this, it is difficult for clinicians to create or edit the contents of existing CIGs, making necessary the involvement of a knowledge engineer for CIG management and implementation [4]. In order to solve this issue, Authoring Tools (AT) are proposed as tools for facilitating and actively taking part in the creation of CIGs by clinical domain experts. To avoid the technical encoding, clinicians are provided with user-friendly and interactive interfaces that ease the formalization process [5].

For example, Dunsmuir et al. [6] developed an AT in order to enable anaesthesiologists to include their clinical knowledge in a rule-based CDSS. Their AT allows users to introduce and edit clinical rules, but requires the user to have knowledge about XML files in order to edit the terminology and parameters used by the AT. Furthermore, the format used to encode clinical rules is designed for its use in anaesthesiology, which makes it too specific for its adoption in other clinical domains.

Another example would be in the case of Ali et al. [7], who propose a framework that makes use of HL7 standard and ontologies in order to generate shareable CIGs. Rules are created and edited using an ontology-linked AT, and Arden Syntax is used to represent them. While the constructed rules follow a standard to encode the clinical knowledge, no standard systems for grading clinical evidence are used.

In this work, following the approach by Muro et al. [8], an AT for developing domain independent CIGs is presented. This proposal uses the Decisional Event (DE) structure to store the clinical information regarding the decision-making process. Using this structure it is possible to compute the quality of evidence and strength of recommendation for each of the rules, as described in [9]. As our platform is based on this DE concept, the strength of the recommendations in the CIG can be assessed.

3. Methodology

The proposed AT, within the architecture shown in Figure 1, is composed by:

- (i) a module for interacting with external knowledge models, such as ontologies, to provide the needed clinical domain knowledge during the formalization process,
- (ii) a user-friendly Graphical User Interface (GUI) as the frontend of the system for introducing rules that will compose the final CIG, and
- (iii) a backend, which is responsible for creating the final CIG.

To ease the technical comprehension of this chapter, it is supported by a use case in the management of gestational diabetes.

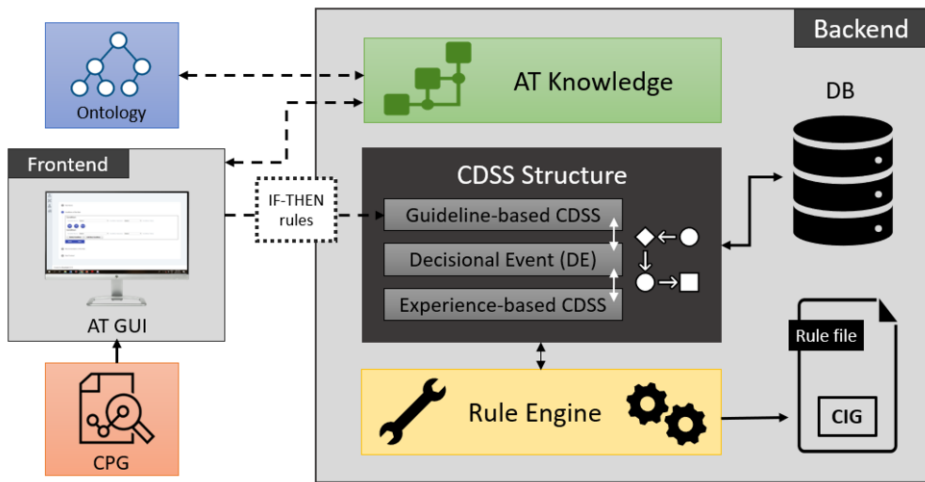


Figure 1. General view of the framework.

3.1. Interaction of the AT with ontologies

Ontologies formalize the clinical knowledge in a standard, flexible and interoperable way. In this approach, ontologies are externally queried to provide the needed clinical domain knowledge during the formalization process. After a research on different APIs for the integration of ontologies with the CDSS, Apache Jena¹ was selected, as it allows our development to integrate ontologies using RDF or OWL models in an easy way.

For the interaction of the AT with the ontology, different web services are used to:

- (i) get the list of variables (classes) from the ontology for defining the variable names within a rule,
- (ii) obtain the possible values of the selected variables, and
- (iii) receive the list of recommendations for completing the consequent part of the rule. This interaction can be seen in Figure 2, where the possible values that a

¹ <https://jena.apache.org/documentation/ontology>

variable (“*CurrentPhysicalActivityLevel*”) can take (left) and a recommendation as the consequent part of a rule (right) are shown.

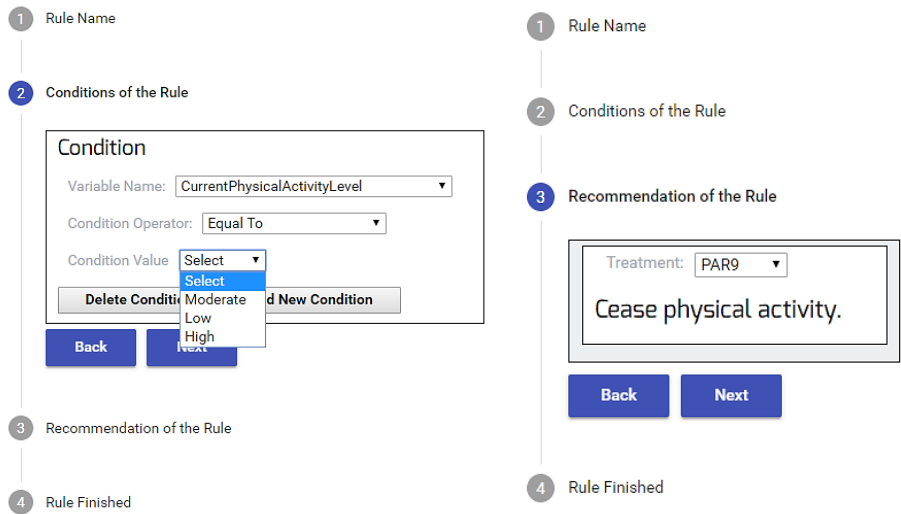


Figure 2. Introduction of conditions (left) and recommendations (right) of rules in the AT.

3.2. CPG formalization

The process of CPG formalization is the result of a continues interaction between the frontend and the backend of the system. The proposed GUI allows the user to formalize rules following four simple steps (see example in Figure 2). First, the name of the rule is defined, which defines the clinical content in a textual way. Afterwards, the conditions of the rule are introduced, defining the statements to be accomplished by the studied patient. Next, the recommendation for the formalized rule is specified as the consequent part to be provided. Finally, the rule is sent to the backend where it is stored and added into the knowledge base of the CDSS to be triggered for new cases study.

The backend supporting this AT is composed by (i) a knowledge layer linked with external knowledge models, such as ontologies, to provide the needed clinical domain knowledge during the formalization process, (ii) a structure layer that gathers all this clinical knowledge and formalizes it in a technology-independent way relaying in the DE concept explained in [9], (iii) a rule engine that integrates a rule file generator for automatic triggering of the formalized knowledge, and (iv) a database (DB) that stores the formalized rules.

With the interaction between the frontend and the backend of the system, the rules in the formalized CIG can be retrieved for their edition and update. Moreover, the used DE structure allows to write down CIGs in any document-based format (e.g. .drl, .xml, .json), since the knowledge is built in a java-based structure [8]. Finally, using the information stored in the DE concept, it is possible to apply standard metrics for

grading clinical evidence such as GRADE² to assess the quality of the evidence included in the guidelines.

4. Conclusions and Future Work

In this paper an Authoring Tool (AT) for the formalization of CPGs into CIGs is presented. This AT allows clinicians to take an active role in this process, as it enables them to formalize CPGs and personalize them according to their own experience. Also, the system facilitates the maintenance and shareability of this knowledge, as the followed DE structure can be exported into other clinical standards. With the use of this DE concept, the clinical evidence of the recommendations obtained from the system can be evaluated. In addition, the proposed system is domain-independent, allowing to formalize CPGs from different domains using the same AT. The generated CIG can then be integrated in a CDSS to provide patient-specific recommendations. Furthermore, new clinical knowledge can be added to the CIG using the AT GUI in order to update its contents with the latest evidence-based clinical practice.

Overall, the presented tool eases the formalization of CPGs into CIGs for clinicians or knowledge engineers, while also allowing to edit and update the introduced knowledge. Furthermore, tools for assessing the reported clinical evidence in the CIGs are provided, avoiding the inclusion of erroneous or low evidence-based knowledge.

Future work will include a visualization of the constructed rules to aid the clinicians' understanding how the introduced knowledge is being formalized. Finally, an ontology editor will be implemented for visualizing its classes and easing the manipulation of the ontology's contents.

References

- [1] Institute of Medicine (US) Committee to Advise the Public Health Service on Clinical Practice Guidelines, *Clinical Practice Guidelines: Directions for a New Program*. Washington (DC): National Academies Press (US), 1990.
- [2] B. Hurwitz, "Legal and political considerations of clinical practice guidelines," *BMJ*, vol. 318, no. 7184, pp. 661–664, Mar. 1999.
- [3] J. Bouaud and B. Seroussi, "Impact of site-specific customizations on physician compliance with guidelines," *Stud Health Technol Inform*, vol. 90, pp. 543–547, 2002.
- [4] A. Seyfang, S. Miksch, M. Marcos, J. Wittenberg, C. Polo-Conde, and K. Rosenbrand, "Bridging the Gap between Informal and Formal Guideline Representations," p. 5, 2006.
- [5] L. Zhou et al., "A study of diverse clinical decision support rule authoring environments and requirements for integration," *BMC Med Inform Decis Mak*, vol. 12, p. 128, 2012.
- [6] D. Dunsmuir, J. Daniels, C. Brouse, S. Ford, and J. M. Ansermino, "A Knowledge Authoring Tool for Clinical Decision Support," *J Clin Monit Comput*, vol. 22, no. 3, p. 189, 2008.
- [7] T. Ali, M. Hussain, W. Ali Khan, M. Afzal, and Sungyoung Lee, "Authoring tool: Acquiring sharable knowledge for Smart CDSS," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, pp. 1278–1281, 2013.
- [8] N. Muro et al., "Architecture for a Multimodal and Domain-Independent Clinical Decision Support System Software Development Kit," 2019.
- [9] N. Muro et al., "Augmenting Guideline Knowledge with Non-compliant Clinical Decisions: Experience-Based Decision Support," in *Innovation in Medicine and Healthcare 2017*, vol. 71, Cham: Springer International Publishing, pp. 217–226, 2018.

² <http://www.gradeworkinggroup.org/>

A Process Mining Application for the Analysis of Hospital-at-Home Admissions

Ilaria Angela AMANTEA^a, Emilio SULIS^{a,1}, Guido BOELLA^a,
Renata MARINELLO^b, Dario BIANCA^b, Enrico BRUNETTI^b, Mario BO^b and
Carlos FERNANDEZ-LLATAS^c

^a *Computer Science Department, University of Turin, 185 c. Svizzera, 10149 Turin, Italy*

^b *City of Health and Science, 88 c. Bramante, 10126 Turin, Italy*

^c *SABIEN-ITACA-Universitat Politècnica de València, Camino de Vera S/N, 46022 Valencia, Spain*

Abstract. This article proposes the analysis of the admissions to hospital-at-home service within the framework of process mining. In addition to conventional modeling in standard languages, relying on interviews and continuous improvement, we propose the adoption of an automatic process discovery technique based on data collected by the hospital information system. We focus on the patient admission process, in which staff discriminate cases of interest for the service. Our methodological framework starts with the extraction of process information from the existing dataset. Once obtained meaningful data for an event log analysis, we propose the adoption of a process discovery algorithm by using a specific tool for process mining. In the context of Business Process Management, we suggest a practical application to be explored in order to improve standard modeling, opening the way to perform business process simulation with scenario analysis.

Keywords. Business Process Management, Process Mining, E-health, Hospital-at-Home

1. Introduction

In Medicine, one of the important aspects to investigate is the organization of health processes, by considering an holistic and systemic approach. In this direction, computer science gained a relevant role for technical hardware and information systems improvements [8]. This is the case of Business Process Management (BPM) [7], a discipline combining data science and management studies to perform business process improvement, modeling and simulation [19] for decision-making [2], also in healthcare [12,15]. Modeling usually adopt standard languages, e.g. BPMN², as well as relevant data collected from Process-Aware Information Systems [8], properly stored with the recent XES format [3]. Such workflow analysis facilitates the detection of inefficiencies, bottlenecks, constraints, and risks [20,5,18].

In this paper, we explore the adoption of a process discovery technique from Process Mining [9], a recent and promising field of study, in order to automatically extract

¹Corresponding Author, email: emilio.sulis@unito.it

²See OMG, Business Process Model Notation: <https://www.omg.org/spec/BPMN/2.0/About-BPMN/>

relevant information from event log data concerning the patients workflow [1]. We focus on a specific healthcare service, the Hospital-at-Home service (HaH) in Torino, one of the largest city of Italy, to investigate the admission process³.

The availability of real data about the service is at the core of healthcare business process mining perspective [13,16]. Nevertheless, the adoption of Process Aware Information System (PAIS) in most hospitals is not yet achieved, and healthcare managers more often deal with conventional aggregated information. A large part of hospital information systems actually is not well equipped for process analysis, whereas data can be difficult to extract, including several well-know problem concerning data quality [14].

This work focuses on two research problems: i. Can we derive digital information even in the absence of an information system that accurately collects data? ii. On this basis, can we automatically derive an healthcare process, e.g. the HaH admissions? We improve similar works which addressed the specific topic of modeling home-care services [17], by applying process discovery technique to data of the service.

In the following of the paper, we introduce our use case, dataset and the methodological framework in Section 2, while Section 3 includes the output of process mining. We draw some conclusions in Section 4.

2. Use case, data and methodology

The scientific literature as well as the practical experiences have already recognized the usefulness of domiciliary cares, pushing hospitals to operate with integrated territorial services [21].

2.1. Description of the use case

In the context of HaH carried out since 1985 by the main hospital in Torino, one of the biggest city in northern Italy, we focus on the experimental project of geriatric home hospitalization, in Piedmont region. This service started in the Molinette Hospital within the “City of Health and Science” of Torino, which actually is one of the largest public health hub in Europe⁴. This innovative service ensures cares mostly for elderly and fragile people in acute disease. In addition to institutionalized services for patients, HaH provides a relevant support to caregivers and their relatives. This avoids the improper accesses to Emergency services, contributing in the reduction of the consistency of waiting lists. The main interest of hospital managers to investigate HaH by adopting a process-oriented perspective relies in the evaluation of benefits to further extend this kind of service.

2.2. Business process modeling and data analysis

In Figure 1 we provide a brief overview of HaH service, focusing on admission process, whose diagram is on the left part. Eligible patients for the service are interviewed by a Case Manager (CM), to investigate and evaluate the existing conditions to apply for HaH.

³This research is part of the project “CANP - CAa Nel Parco” of Regione Piemonte funded by POR FESR PIEMONTE 2014-2020. Web site of the project: <http://casanelparco-project.it>.

⁴Cfr. City of Health and Science, <http://www.cittadellasalute.to.it>, as well as: <http://shorturl.at/bosz3>

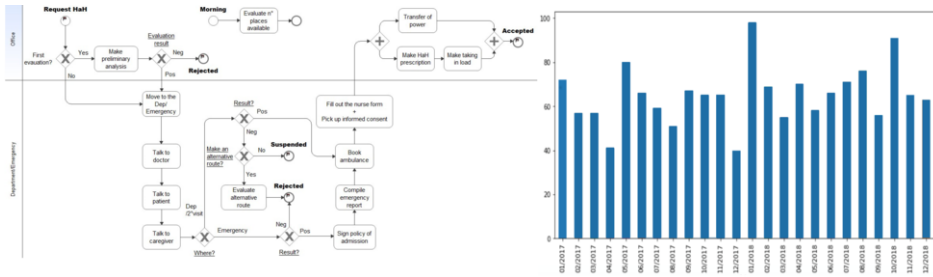


Figure 1. On the left: business process analysis in BPMN. On the right: Arrivals of patients at the HaH service in 2017 and 2018 in the different months.

The diagram describes the output of a conventional approach with BPMN. Due to lack of relevance in this paper, we avoid to detail all the tasks and passages, better presented in [4]. In addition, we perform traditional data analysis to contextualize the process of interest. The histogram in the right side of Figure 1 summarizes the arrivals of patients in last two years by months, with peaks corresponding to the periods of flu (January and Autumn), and less requests near Italian festivity days (i.e. April for Easter festivities).

2.3. Log of the process

To promote a business process analysis, we need more information about the patients' flow. An initial step in the construction of the admission process includes the exploitation of data registered in the hospital information system. As in the log example presented in the following paragraph, we extracted information including: patient's number (ID), corresponding case (SDO), provenience (PROV), first call to the CM for the evaluation (CALL), admission date (HaH1), type (OUT) and date of dismissal (END).

ID, SDO, PROV, CALL, HaH1, OUT, END

045, 2017500110, SPEC, 02/02/17, 06/02/17, ORD, 14/02/17

048, 002017500552, PS, 21/06/17, 27/06/17, TRASF, 09/07/17

045, 2018500025, MDB, 04/01/18, 04/01/18, DEC, 25/02/18

2.4. Process Mining in Health

To perform the discovery of the process from the log we adopted PALIA Process Discovery algorithm [10]. In particular, we apply PALIA algorithm for discovering the process of HaH events described in previous section. PALIA Algorithm has been successfully used in several medical problems like surgery [10], Emergency Rooms [11], or Diabetes [6], among other health solutions. Process Mining can support healthcare professionals in the real understanding of the process, not only by discovering the process but also performing assessments for enabling the application of Value based healthcare [11].

3. Results

The result of our process mining technique is the model presented in Figure 2, whereas the admission process has been automatically extracted from raw data.

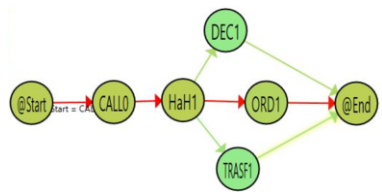


Figure 2. Output of the admission process

The process starts with the patient admission (HaH1) preceded by a call by the Case Manager (CALL0), and followed by three types of output: the ordinary management of HaH activities (ORD) until the dismissal (END), the transfer of the patient - e.g. surgery or the recovery to an Hospital department (TRAS1); or the patient’s death (DEC).

A second model is presented in Figure 3, describing also the amount of occurrences. Here, the color reflect the frequency of each event: as occurrences increase, arrows and circle tend to red color. As clearly stated by the figure, the most frequent output is patients hospitalization (only once time) and then the ordinary discharged.

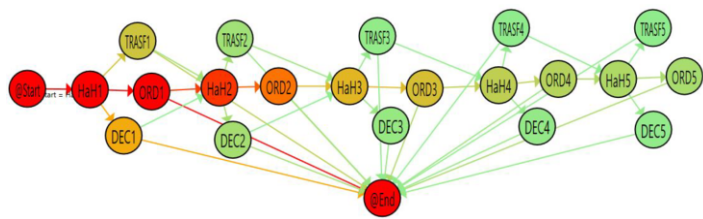


Figure 3. Process discovery from HaH admission dataset.

By investigating our dataset, we immediately observe how the service performs well, in general, as most patients are discharged in an ordinary way. In fact, a further investigation identifies how most patients, once discharged, do not return to the hospital. In addition, it can be observed that the number of re-hospitalization is not very relevant.

4. Conclusions

We applied process mining to data about the admission of patients in healthcare service. One point of interest is to demonstrate the ability to perform a healthcare process analysis, by automatically deriving information about processes from not well structured data (i.e., data not registered in standard event log XES format), as actually most hospital information systems are still not process-aware. We believe that the understanding of the advantages of using process mining techniques will push towards a wider adoption of PAIS in the near future, as it is actually happening in Industry. As a matter of fact, the methodological framework here presented allows to compare traditional process analysis with the emerging techniques of process discovery, opening the way to perform conformance checking and enhancement.

References

- [1] Van der Aalst, W.: *Process mining: discovery, conformance and enhancement of business processes*, vol. 2. Springer (2011)
- [2] Van der Aalst, W., Nakatumba, J., Rozinat, A., Russell, N.: *Business process simulation*. In: *Handbook on BPM 1*, pp. 313–338. Springer (2010)
- [3] Acampora, G., Vitiello, A., Di Stefano, B., Aalst, W.M.P.v.d., Günther, C.W., Verbeek, H.M.W.: IEEE 1849TM: The XES Standard. *IEEE Computational Intelligence Magazine* pp. 4–8 (May 2017)
- [4] Amantea, I., Arnone, M., Di Leva, A., Sulis, E., Bianca, D., Brunetti, E., Marinello, R.: Modeling and simulation of the hospital-at-home service admission process. In: *Proceedings of the 9th International Conference on Simulation and Modeling Methodologies, Technologies and Applications*. pp. 293–300. SCITEPRESS-Science and Technology Publications, Lda (2019)
- [5] Amantea, I.A., Leva, A.D., Sulis, E.: A simulation-driven approach in risk-aware business process management: A case study in healthcare. In: *Proceedings of SIMULTECH*. pp. 98–105. INSTICC, SciTePress (2018). <https://doi.org/10.5220/0006842100980105>
- [6] Conca, T., Saint-Pierre, C., Herskovic, V., Sepúlveda, M., Capurro, D., Prieto, F., Fernandez-Llatas, C.: Multidisciplinary collaboration in the treatment of patients with type 2 diabetes in primary care: Analysis using process mining. *Journal of medical Internet research* **20**(4) (2018)
- [7] Dumas, M., La Rosa, M., Mendling, J., Reijers, H.: *Fundamentals of business process management*, vol. 1. Springer, 2nd edn. (2018)
- [8] Dumas, M., Van der Aalst, W.M., Ter Hofstede, A.H.: *Process-aware information systems: bridging people and software through process technology*. John Wiley & Sons (2005)
- [9] Fernández-Llatas, C., García-Gómez, J.M.: *Data mining in clinical medicine*. Springer (2015)
- [10] Fernandez-Llatas, C., Lizondo, A., Monton, E., Benedi, J.M., Traver, V.: Process mining methodology for health process tracking using real-time indoor location systems. *Sensors* **15**(12), 29821–29840 (Nov 2015)
- [11] Ibanez-Sanchez, G., Fernandez-Llatas, C., Celda, A., Mandingorra, J., Aparici-Tortajada, L., Martinez-Millana, A., Munoz-Gama, J., Sepúlveda, M., Rojas, E., Gálvez, V., Capurro, D., Traver, V.: Toward value-based healthcare through interactive process mining in emergency rooms: The stroke case. *International Journal of Environmental Research and Public Health* **16**(10) (2019)
- [12] Johnson, O.A., Dhafari, T.B., Kurniati, A., Fox, F., Rojas, E.: The clearpath method for care pathway process mining and simulation. In: *International Conference on Business Process Management*. pp. 239–250. Springer (2018)
- [13] Mans, R., Van der Aalst, W., Vanwersch, R.: *Process mining in healthcare: evaluating and exploiting operational healthcare processes*. Springer (2015)
- [14] Martin, N., Depaire, B., Caris, A.: The use of process mining in business process simulation model construction. *Business & Information Systems Engineering* **58**(1), 73–87 (2016)
- [15] Pegoraro, F., Santos, E.A.P., Loures, E.d.F.R., da Silva Dias, G., dos Santos, L.M., Coelho, R.O.: Short-term simulation in healthcare management with support of the process mining. In: *World Conference on Information Systems and Technologies*. pp. 724–735. Springer (2018)
- [16] Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D.: Process mining in healthcare: A literature review. *Journal of biomedical informatics* **61**, 224–236 (2016)
- [17] Russo, V., Ciampi, M., Esposito, M.: A business process model for integrated home care. In: *EU-SPN/ICTH* (2015)
- [18] Sulis, E., Amantea, I.A., Boella, G., Marinello, R., Bianca, D., Brunetti, E., Bo, M., Bianco, A., Cattel, F., Cena, C., Fruttero, R., Traina, S., Feletti, L.C., de Cosmo, P., Armando, L., Ambrosini, S.: Monitoring patients with fragilities in the context of de-hospitalization services: An ambient assisted living healthcare framework for e-health applications. In: *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*. pp. 216–219 (June 2019). <https://doi.org/10.1109/ISCE.2019.8900989>
- [19] Sulis, E., Di Leva, A.: An agent-based model of a business process: The use case of a hospital emergency department. In: *Business Process Management Workshops. Lecture Notes in Business Information Processing*, vol. 308, pp. 124–132. Springer (2017)
- [20] Suriadi, S., Weiß, B., Winkelman, A., ter Hofstede, A.H., Adams, M., Conforti, R., Fidge, C., La Rosa, M., Ouyang, C., Rosemann, M., et al.: Current research in risk-aware business process management: overview, comparison, and gap analysis. *Communications of the AIS* **34**(1), 933–984 (2014)
- [21] Tibaldi, V., Ricauda, N.A., Rocco, M., Bertone, P., Fanton, G., Isaia, G.: L'innovazione tecnologica e l'ospedalizzazione a domicilio. *Recenti Progressi in Medicina* **104**(5), 181–188 (2013)

An Albanian Text-to-Speech System for the BabelDr Medical Speech Translator

Nikos TSOURAKIS^{a,1}, Rovena TROQE^b, Johanna GERLACH^a, Pierrette BOUILLON^a and Hervé SPECHBACH^c

^aUniversity of Geneva, Switzerland

^bUniversity of the Free State, Bloemfontein, South Africa

^cGeneva University Hospitals, Switzerland

Abstract. In this paper we present work on creating and evaluating a Text-to-Speech system for the Albanian language to be used in the BabelDr medical speech translation system. Its quality was assessed by twelve native speakers who provided feedback on 60 prompts generated by the synthesizer and on 60 real human recordings across three dimensions, namely comprehensibility, naturalness and likeability. The results suggest that the newly created voice can be incorporated in the content creation pipeline of the BabelDr platform.

Keywords. Text-to-Speech, Albanian, Tacotron 2, BabelDr

1. Introduction

The ever increasing movement of people worldwide poses new challenges for healthcare institutions across Europe in particular problems related to different language and cultural barriers experienced by the healthcare practitioners and their foreign patients [1], [2]. Most of the time experienced interpreters, relatives or friends try to bridge this communication gap. However, this solution is often less than satisfactory as trained medical interpreters are both scarce and expensive, family members might not communicate the exact meaning of the physician's question or the patient's response, while there are also privacy concerns. It is generally agreed that translation quality is of prime importance in this setting [3], while tools like Google Translate pose certain problems on reliability and data privacy.

Since 2017 the Faculty of Translation and Interpreting of the University of Geneva and the Geneva University Hospitals (HUG) joined forces for building a medical speech translator, under the BabelDr project² funded by the Fondation Privée des HUG. The system is a web based tool that has been specifically designed to assist in triaging of non-French-speaking patients visiting HUG's A&E department, and allows a medical

¹ Nikos Tsourakis, TIM/FTI, University of Geneva, 40, boulevard du Pont-d'Arve 1211 GENEVE 4 Switzerland; E-mail: Nikolaos.Tsourakis@unige.ch.

² <https://babeldr.unige.ch>

³ <https://forms.gle/HZoJszhmKgbmeJqYA>

⁴ <https://github.com/mozilla/TTS>

⁵ <https://github.com/erogol/WaveRNN>

⁶ <https://github.com/NVIDIA/waveglow>

professional to perform a preliminary medical examination dialogue to determine the nature of the patient's problem and the appropriate action to take. The system incorporates state-of-the-art technologies like speech recognition (ASR) and text-to-speech (TTS) and supports translation for different language pairs [4]. It is also aligned with mainstream research findings [5] on the basis that it incorporates preset health related questions and declarative sentences (11.134 in total) that can be selected using around 1M variations of similar speech phrases.

In the context of this work we created and evaluated a synthetic female voice for the Albanian language based on Tacotron 2, a neural network architecture for speech synthesis directly from text [6] that can be used for announcing the physician's questions to an Albanian speaking patient. In this paper we present the data gathering process and the steps taken in order to train the TTS. Based on an evaluation with twelve native speakers we assessed the comprehensibility, naturalness and likeability of the newly created voice. According to our findings we can reasonably argue that the synthesized voice can be incorporated in the content creation pipeline of the BabelDr platform.

Section 2 presents the methods used for gathering audio data, training the voice and evaluating its quality. Section 3 reports on the results of the training process and the feedback from the native speakers. Section 4 discusses the findings and different implications in more depth. The final section concludes.

2. Method

2.1. TTS training

As in every training task that involves a neural network architecture the amount of data is a crucial factor. Using an in-house web-based recording tool and an intermediate fidelity microphone we asked a female subject to perform 10452 recordings from home using a sampling rate of 48 kHz. The text included questions from 13 diagnostic domains with phrases like: "Can you show me with the finger where the pain is?", "Are you allergic to antibiotics?", "How many cigarettes do you smoke per day?", etc. The created corpus contained 2014 unique words, while the average length of the sentences was 14 words (sd=6). The corpus was split into a training set consisting of 9500 sentences and an evaluation set with 952 sentences. Approximately 9.5 hours of speech were recorded.

2.2. Human evaluation

Contrary to other evaluation tasks, where software systems can be assessed in terms of various objective measurements, evaluating a synthesized voice is highly subjective. For this reason we recruited twelve native Albanian speakers and asked them to provide their feedback across three dimensions typically encountered in TTS evaluations [7]. Specifically, they had to listen to a series of audio prompts and express their opinion in a five-point Likert scale (subjective mean opinion score - MOS) concerning how comprehensible the prompt is, how natural the voice sounds and how much they like the voice. The whole evaluation was set-up as a Google form³ so that the participants could provide their feedback easily, while the investigators of the study could acquire their responses immediately. The evaluation protocol was based on two hypotheses:

- The TTS can be used in our content creation pipeline.

- The TTS works equally well for short and long sentences.

For testing the two hypotheses we generated 60 audio prompts with the TTS and selected randomly 60 volume normalized recordings from the initial training corpus. For each of these two subsets we chose sentences of different word length according to this formula: 20 sentences with 1 to 3 words, 20 sentences with 4 to 9 words and 20 sentences with more than 9 words. The human recordings and the TTS prompts were merged and shuffled into a single set, so that participants were not aware of the source of each one.

3. Results

3.1. TTS training

The TTS engine was trained based on a Tacotron 2 implementation⁴ using the following hyper-parameters: learning rate=0.0001, batch size=32 and epochs=1000. The model had to learn ~28M parameters and the process took place on a Linux server equipped with an NVIDIA GeForce GTX Titan X, 12GB GPU card. Figure 1 shows the average total loss (average of the linear loss, mel loss and stop loss) in respect to the training epochs, which reaches to 0.01 at the last one.

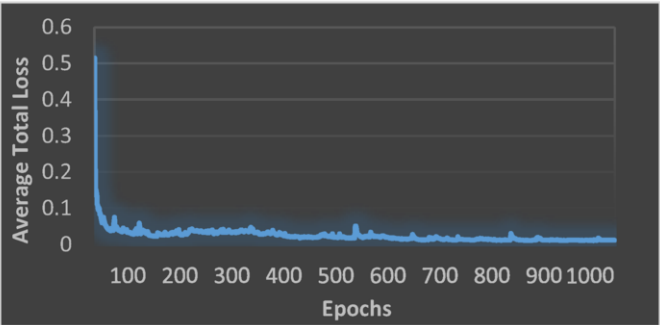


Figure 1. Average total loss per number of epochs

3.2. Human evaluation

Table 1. Mean opinion score (MOS) for the three evaluation criteria: Comprehensibility (C), Naturalness (N) & Likeability (L) and for the two types of prompts

	I understand the question (C)	The voice sounds natural (N)	I like the voice (L)
TTS MOS	3.9 (sd=0.5)	3.0 (sd=1.0)	3.1 (sd=1.1)
Recordings MOS	4.5 (sd=0.4)	4.2 (sd=0.9)	4.1 (sd=0.9)

Table 1 summarizes the mean opinion score of all participants, for the three evaluation criteria and the two types of prompts. As expected human recordings outperform the TTS in all dimensions. Although the TTS might sound less natural and be less likeable it is still quite comprehensible 3.9/5 vs. 4.5/5 (statistically significant difference at $p<0.002$). We also examine the variation in comprehensibility, naturalness and likeability of the TTS based on the sentence length. As it can be observed in Table 2, our initial hypothesis

is correct and the TTS works equally well for short and long sentences (no statistically significant differences).

Table 2. Mean opinion score (MOS) of Comprehensibility (C), Naturalness (N) & Likeability (L) of the TTS prompts based on the sentence word length

	MOS (C)	MOS (N)	MOS (L)
Sentences with 1-3 words	3.8 (sd=0.6)	3.0 (sd=1.0)	3.1 (sd=1.1)
Sentences with 4-9 words	4.0 (sd=0.4)	3.1 (sd=1.0)	3.1 (sd=1.1)
Sentences with more than 9 words	3.9 (sd=0.6)	3.0 (sd=1.0)	3.0 (sd=1.1)

3.3. Qualitative analysis

In order to qualitatively analyze the source of errors in the synthesized prompts we asked two experts to identify possible problems. Specifically, each expert had to assess each prompt based on possible intonation problems (e.g. not being able to distinguish if it is a question or a request) and whether the prompt included muffled or unnatural sounds. These criteria were selected on the basis of informal feedback by the participants of the evaluation task. We calculated the Cohen’s Kappa score for each one of the two pairs that expresses agreement corrected for chance, shown in Table 3. A kappa ranges between -1 and 1 and a value of 0.19 is defined as slight agreement whereas a value of -0.27 as a poor one [8]. The last row shows the absolute number of sentences where the two experts agreed.

Table 3. Cohen’s Kappa for the two kinds of possible problems (95% confidence intervals in parenthesis) and absolute sentence agreement

	Intonation problems	Muffled or unnatural sounds
Cohen’s Kappa	0.19 (0.1, 0.34)	-0.27 (-0.51, -0.04)
Agreement	30 out of 60	25 out of 60

We also calculated the correlation between the MOS of comprehensibility (continuous variable) and the responses from each expert and for each one of the two criteria (categorical variables “0” or “1”). The appropriate method in this case is the point biserial correlation, which did not show any significant differences. Both correlation and agreement suggest that using intonation and checking for unnatural sounds is not an appropriate method for assessing prompt comprehensibility, which should be determined on a case-by-case basis with other criteria.

4. Discussion

In the era of neural networks creating a new voice from scratch demands few human months of work something that a decade ago would require extensive expertise and commitment for much longer time. After the positive experience with the Albanian TTS we are capable of creating new voices according to the needs of the project. Resorting to a commercial TTS is not always an option due to the lack of support for certain languages; a typical example is Tigrinya, frequently encountered at HUG, for which no synthetic voice yet exists. Even for those languages where a commercial option is available it normally targets the most common dialect. Arabic is another typical example.

On the other hand one might argue in favor of recording every new sentence from scratch. This option is well aligned with the BabelDr project which is strongly biased

towards quality assurance. Every translation is checked by experts to avoid possible errors and ambiguities. The sole difference, however, is that multiple people can perform the translation task but a single person can do the audio recordings. If this person is temporary not available the whole deployment pipeline is stalled. Even as a backup strategy having your own TTS presents many competitive advantages. Nonetheless there is always a human in the loop in our content creation pipeline that checks both translations and TTS prompts and decides if they are acceptable or not.

A factor that was not properly addressed in this study concerns on how we measure comprehensibility. In essence by asking participants to quantify their understanding you risk obtaining a negative answer for a perfectly oralized prompt simply because the question was ambiguous or unclear in the first place (e.g. a medical term was used). In a future study two levels of comprehensibility should be scrutinized by carefully creating and testing a corpus that alleviates this deficiency. Conversely, even a perfect score in comprehensibility does not decisively mean that the prompt sounds perfect. The participants expressed concerns that sometimes they listened to sentences that did not sound completely right but were still comprehensible.

5. Conclusion

Language barriers often cause inconvenience but when medical issues are involved they cease to be mere inconvenience and can become life-threatening. In this work we include a new building block in our platform for creating synthesized voices. After evaluating its feasibility for the Albanian language with twelve native speakers we can include it in the BabelDr content creation pipeline. We have also found that there are no differences in quality between short and long sentences.

In the future we plan to combine Tacotron 2 with a neural Vocoder like WaveNet⁵ and WaveGlow⁶ that promise to ameliorate the output quality. We also intend to gather more data for Albanian, as 9.5 hours of training material puts us in the low end of complexity (e.g. in [6] around 25 hours of speech were used). We will also create a new voice for the Tigrinya language.

References

- [1] Flores, G., Laws, M.B., Mayo, S.J., Zuckerman, B., Abreu, M., Medina, L. and Hardt, E.J. Errors in medical interpretation and their potential clinical consequences in pediatric encounters. *Pediatrics*, **111**(1), (2003).
- [2] Wasserman, M., Renfrew, M.R., Green, A.R., Lopez, L., Tan-McGrory, A., Brach, C. and Betancourt, J.R. Identifying and preventing medical errors in patients with limited English proficiency: key findings and tools for the field. *Journal for Healthcare Quality*, **36**(3), 2014, 5-16.
- [3] Tsourakis, N. and Estrella, P. Evaluating the quality of mobile medical speech translators based on ISO/IEC 9126 series: definition, weighted quality model and metrics. *IJRQEH*, **2**(2), 2013, 1–20.
- [4] Spechbach, H., Gerlach, J. Karker, S.M., Tsourakis, N., Combescure, C. and Bouillon, P. A Speech-Enabled Fixed-Phrase Translator for Emergency Settings: Crossover Study. *JMIR Medical Informatics*, **7**(2) (2019).
- [5] Panayiotou, A., Gardner, A., Williams, S., Zucchi, E., Mascitti-Meuter, M., Goh, A.M., You, E., et al. Language Translation Apps in Health Care Settings: Expert Opinion. *JMIR Mhealth Uhealth*, **7**(4), 2019.
- [6] Shen, J., Pang, R., Weiss, R.J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., et al. Natural TTS synthesis by conditioning Wavenet on mel spectrogram predictions. *IEEE ICASSP*, (2018), 4779-4783.
- [7] Dybkjaer, L., Hemsén, H. and Minker, W. *Evaluation of Text and Speech Systems (1st ed.)*. Springer, 2007.
- [8] Landis, J.R. and Koch G.G. The measurement of observer agreement for categorical data. *Biometrics*, (33), 1977, 159–174.

Are Clinical Decision Support Systems Compatible with Patient-Centred Care?

Vije Kumar RAJPUT^a, Jack DOWIE^{b,c1}, Mette Kjer KALTOFT^c

^a*Stonydelph Health Centre, Tamworth, UK*

^b*London School of Hygiene and Tropical Medicine*

^c*University of Southern Denmark*

Abstract. Few, if any, of the Clinical Decision Support Systems developed and reported within the informatics literature incorporate patient preferences in the formal and quantitatively analytic way adopted for evidence. Preferences are assumed to be ‘taken into account’ by the clinician in the associated clinical encounter. Many CDSS produce management recommendations on the basis of embedded algorithms or expert rules. These are often focused on a single criterion, and the preference trade-offs involved have no empirical basis outside an expert panel. After illustrating these points with the Osteoporosis Adviser CDSS from Iceland, we review an ambitious attempt to address both the monocriterial bias and lack of empirical preference-sensitivity, in the context of Early Rheumatoid Arthritis. It brings together the preference data from a Discrete Choice Experiment and the best available evidence data, to arrive at the percentage of patients who would prefer particular treatments from those in the listed options. It is suggested that these percentages could assist a GRADE panel determine whether to produce a strong or weak recommendation. However, any such group average preference-based recommendations are arguably in breach of both the reasonable patient legal standard for informed consent and simple ethical principles. The answer is not to localise, but personalise, decisions through the use of preference-sensitive multi-criteria decision support tools engaged with at the point of care.

Keywords: Clinical Decision Support System, guidelines, GRADE, multi-criteria decision support, osteoporosis, rheumatoid arthritis

1. Introduction

Clinical Decision Support Systems (CDSS) can approach the eponymous task in a variety of ways, but this variation is within the taken-for-granted assumption, understandable in the context of ‘informatics’, that the basic support for the clinical decision at the point of care is in the form of information. However, it is clear that this excludes information on patient preferences. The aim here is to question this practice and argue for a rethinking of CDSS that would bring them into line with recent moves toward patient- and person-centred care. To avoid any suggestion of misrepresentation we start with a mainstream definition [1], with which all the CDSS presented, as such at MIE and Medinfo conferences in the last five years are in line.

¹ Corresponding author: Jack Dowie, LSHTM, 15-17 Tavistock Place, London, UK WC1H 9SH; email: jack.dowie@lshtm.ac.uk

Clinical decision support systems (CDSS) - defined as any system designed to improve clinical decision-making related to diagnostic or therapeutic processes of care... use specific parameters (such as diagnoses, laboratory results, medication choices, or complex combinations of clinical data) to provide information or recommendations directly relevant to a specific patient encounter at the point of care... CDSS address activities ranging from the selection of medications (e.g., the optimal antibiotic choice given specific microbiologic data) or diagnostic tests (e.g., the best blood test to evaluate a patient with possible pulmonary embolism) to detailed support for optimal drug dosing and support for resolving diagnostic dilemmas... Typical CDSS suggest default values for drug doses, routes of administration, and frequency, and offer more sophisticated drug safety features such as checking for drug allergies or drug-drug interactions[1].

The AHRQ summary goes on to suggest that the advent of advanced analytic methodologies for large, complex data sets and the development of machine learning techniques is likely to lead to the development of increasingly powerful and sophisticated CDSS. 'These artificial intelligence approaches have tremendous potential for transforming diagnosis and therapy.... However... acceptance of artificial intelligence-based CDSS will require transparency around the methods used to develop recommendations...'. Otherwise the spectre of 'digital paternalism is raised.

Crucially, there is no reference to preferences, which are necessary in any 'selection', 'optimization', or 'recommendation' task. Confirming the basic point of this paper, *Patient-centered* clinical decision support systems (PCCDS) are regarded by AHRQ as a separate category of CDSS: 'PCCDS refers to decision support systems that support individual patients, caregivers, and health care teams in health-related decisions and actions by leveraging patient-specific information (e.g., patient-generated health data) and patient-centered outcomes research findings...'. However, even here, the patient is only as a source of biomedically relevant information, not of preferences.

We employ a simple three-level taxonomy. A *basic CDSS* confines itself to producing biomedical information to assist the clinician to arrive at the best possible diagnostic judgment and therapeutic decision recommendation for this patient, *containing no further guidance or assistance*. This information can range from a single piece of data on a single parameter, such as this patient's bone mineral density (BMD) as measured by a DXA scan, to the output of a multi-attribute model yielding a probability for this patient experiencing a hip fracture in the next ten years (e.g. Frax). A *diagnostic CDSS* goes further, applying embedded *guidelines and/or algorithms to this data* to help the clinician arrive at the best possible diagnostic judgment for this patient - such as whether they meet the criteria for a diagnosis of osteoporosis or osteopenia according to the BMD result. Finally, a *therapeutic CDSS* applies embedded *guidelines and/or algorithms to the basic data and/or diagnostic suggestion/s* to help the clinician arrive at the best possible therapeutic decision recommendation for this patient - such as applying the National Osteoporosis Guideline Group's Frax thresholds for reassuring, treating without testing, or treating and testing. ('Diagnostic judgment' covers the entire range from a single diagnosis to a probabilistic differential diagnosis. 'Therapeutic decision recommendation' refers to any test and/or treatment option).

2. Osteoporosis Advisor: A conventional CDSS

We have already used bone health and fragility fractures as examples and now explore Osteoporosis Advisor (OPAD), an Icelandic commercial product distributed through professional providers and exclusively for use in association with a clinical encounter [2,3]. To document what is, and what is not, in OPAD, we quote at length:

... the Osteoporosis Advisor or OPAD ([http:// www.expeda.is](http://www.expeda.is)), a computerized clinical decision support system to assess and treat osteoporosis in a primary care setting... OPAD provides a patient's 10-year fracture probability (with or without BMD values) using a speedometer-like output that users can easily understand. The OPAD also notes whether a patient would benefit from BMD measurement by DXA. OPAD provides patients with lifestyle and treatment recommendations to reduce their fracture risk, incorporating country-specific guidelines for therapy... OPAD not only gives the fracture risk, but also recommend clinical decisions based on guidelines... Further patient factors may be incorporated into the risk assessment model of OPAD, such as loss of height, history of falls and it is even possible to add individual genotypes to improve the fracture risk estimation [2] (pp. 2, 6, 7).

Notably these 'further patient factors' do not include their preferences. The word 'preference' does not appear in this paper or the earlier companion one. As in most therapeutic CDSS, the patient is a biological object to be subjected to a biomedical assessment, followed by the application of guidelines incorporating either population-based thresholds (NOGG) or, as in OPAD, local expert-based rules (unpublished).

It seems fair to characterise OPAD as a good example of the 'digital' (i.e. data-driven) paternalism' that underlies many therapeutic CDSS.

The busy clinician may have difficulties in interpreting the risk value figure for each patient in hectic daily clinical practice. With this in mind we have extended the information provided by OPAD, by giving a specific diagnosis, that is, osteoporosis or osteopenia, and specific recommendations on prevention, time of next DXA, and treatment options according to international guidelines and experts knowledge... The clinically relevant parameters and the computed risk for fracture are used as input into an expert system which gives specific recommendations for each case with respect to lifestyle changes and treatment options... the system predominantly relies on knowledge capture process of the expert panels, as guidelines never cover all cases. In the end a total of fifteen different treatment recommendations were initially identified as possible recommended treatment options for osteoporosis. The recommended treatments ranged from no treatment to specific recommendations for which drug class was the most appropriate, either as a preventive measure or as a treatment for manifest osteoporosis... [3] (pp. 5,2,3).

There is nothing in any of these statements that would preclude the explicit mention of patients', or patient's, preferences, but there is none. The taken-for-granted assumption is that the latter can - and by implication should be - left to be 'taken into account' by the *clinician*, subsequent to the *clinician's* engagement with the CDSS. Why do CDSS feel justified in producing therapeutic decision recommendations without any reference to information on patient values and preferences?

One answer is that, in line with the embedded algorithm or guideline, many CDSS assume the decision is monocriterial not multicriterial. OPAD assumes that the only criterion that matters is fragility fracture prevention, ignoring all other outcomes (and process considerations) involved in the options. The side effects and treatment burden of the various options are two criteria given significant weight by many patients. In the

piloting of a Danish decision support tool the weights given to avoiding fragility fracture in the next ten years was 43%, to avoiding (option?) side effects 37%, and to avoiding bother/burden 20% [4]. Other populations, settings, and methodologies are likely to produce different results, but highly unlikely to produce anywhere near 100% weight to fragility fracture prevention. It is not a convincing response that these criteria are also to be dealt with outside the CDSS. If a CDSS does not attempt to address a decision as the multi-criterial one it is - probably using some version of Multi-Criteria Decision Analysis [5] - it would be more appropriately called a Clinician Information Support System.

The second answer is that, even absent 'digital paternalism', it is extremely difficult to deal with preferences 'scientifically', since this requires (a) their elicitation and (b) the integration of the results into the decision or guideline. It is not surprising, then, that only in 2019 have we seen a paper attempting it [6].

3. A multi-criterial, group preference-sensitive CDSS

Hazlewood and colleagues accepted the challenge, acknowledging that 'Research evidence alone cannot tell us which treatment is best for patients, and the reason is simple. Even where high-quality evidence is available, treatment choices inevitably involve trade-offs between benefits, risks, dosing, or other monitoring requirements. Incorporating patients' preferences into treatment recommendations has been viewed as the next step in guideline development, but rarely occurs in practice' [6] (p57). They argue that guideline panels should view the assessment and incorporation of *patients'* preferences as a critical step in the evidence to decision process and accordingly seek to assist these panels in making the necessary judgments when translating evidence into a recommendation, using empirically derived data.

In their application to Early Rheumatoid Arthritis (ERA) they 'estimated the preferred treatment using patients' preferences measured in a discrete-choice experiment to apply weights to benefit and harm outcomes from a network meta-analysis and other considerations (dosing, rare adverse events)... We applied this to treatment recommendations for ERA where there is disagreement over the preferred treatment. Thus, a secondary aim was to generate evidence that could be used to help inform patient-centered GRADE treatment recommendations for ERA' [6] (pp.56, 57).

Unfortunately, both the analysis and presentation of results confirm that any approach to the clinical decision through guideline recommendations, even if population-preference informed in this way, is not patient-centred. According to the authors, the patient's preferences - and decision aids - are to come into play only if the percentage of patients preferring a therapy is somewhere between 62% and 78%. Most patients would prefer triple therapy as initial treatment (78%) or after an inadequate response to methotrexate (62%). The probability of choosing triple therapy as initial treatment was further from 50% (the point of indifference) for more patients, making our prediction more confident, and suggesting a stronger recommendation could be made. After an inadequate response to methotrexate, the choice was more split, suggesting a decision aid may be helpful... [6] (p56).

Both proposals - that there is a population percentage threshold for a strong recommendation and that it is at 50% - are arguably in breach of both the reasonable patient legal standard for informed consent and simple ethical principles.

Hazlewood et al. accept that preference studies may be very context-specific, so that existing studies may not be relevant to a guideline group's target population. Hence, generating context-specific preference data would require additional effort, and that may not be feasible. Nevertheless 'Guideline developers could... use a central source of high-quality evidence and find or generate their own patient preference data applicable to their population. This would fit well with... the contemporary axiom of 'globalizing evidence, but localizing decisions'' [6] (p.64).

4. Conclusion

The difficulty of generating relevant group preference data is actually a blessing in disguise. The answer is to personalise, not localise, eliminating the irrelevant detour into group average preferences, and going direct to those of the individual patient in the clinic at the point of care. Conventional therapeutic CDSS should be abandoned, and basic and diagnostic CDSS re-labelled - and evaluated as - *Clinician Information Support Systems*. The best way for informaticians to support preference-sensitive clinical decision making is to redirect their efforts to producing the information (option performance ratings on each of the criteria) required in the Multi-Criteria Decision Analysis-based decision support tools essential for greater patient involvement and transparent preference-sensitivity.

References

- [1] US Dept of Health & Human Services, Agency for Healthcare Research and Quality (AHRQ) Patient Safety Primer August 2019 [<https://psnet.ahrq.gov/primers/primer/51/Clinical-Decision-Support-Systems>]
- [2] B. Halldorsson, A. Bjornsson, H. Gudmundsson, E. Birgisson, B. Ludviksson, B. Gudbjornsson, Clinical decision support system for the diagnosis, fracture risks and treatment of osteoporosis, *Computational and Mathematical Methods in Medicine* (2015), 189769.
- [3] H. Gudmundsson, K. Hansen, B. Halldorsson, B. Ludviksson, B. Gudbjornsson, Clinical decision support system for the management of osteoporosis compared to NOGG guidelines and an osteology specialist: a validation pilot study, *BMC Medical Informatics and Decision Making* **19** (2019), 1–8.
- [4] V. Rajput, M. Kaltoft, J. Dowie, Preference-sensitive multi-criterial decision support can avoid the over- testing produced by mono-criterial guidelines: the case of bone health, *BMJ Evidence-Based Medicine* **24** (2019), A26.
- [5] J. Dowie, M. Kaltoft, G. Salkeld, M. Cunich, Towards generic online multicriteria decision support in patient-centred health care, *Health Expectations* **18** (2013), 689–702.
- [6] G. Hazlewood, C. Bombardier, G. Tomlinson, D. Marshall, A Bayesian model that jointly considers comparative effectiveness research and patients' preferences may help inform GRADE recommendations: an application to rheumatoid arthritis treatment recommendations, *Journal of Clinical Epidemiology* **93** (2018), 56–65.

Case-Based Reasoning for Support of the Diagnostics of Cardiovascular Diseases

Zuzana TOCIMÁKOVÁ^a, Ľudmila PUSZTOVÁ^{a1}, Ján PARALIČ^a,
and Dominik PELLA^b

^a*Department of Cybernetic and Artificial Intelligence,
Faculty of Electrical Engineering and Informatics, Technical University of Košice,
Letná 9, 042 00 Košice, Slovakia*

^b*1st Department of Cardiology, East Slovak Institute of Cardiovascular Diseases,
Ondavská 8, 040 11 Košice, Slovakia*

Abstract. In this paper we present a decision support system, which has been designed and implemented on the case-based reasoning principles. Our decision support system is being implemented in tight cooperation with the cardiologist, who represents the main future users of the system. Our system enables its user to find the most similar historical cases to a new patient, suggest the most probable result of the potential coronary angiography examination and also provide various useful visualizations to the cardiologist, who is responsible for the final decision about recommending the coronary angiography or not for the new patient. The first response from the cardiologist about our system is very promising.

Keywords. Case-based reasoning, k -Nearest Neighbors, Medical diagnostics

1. Introduction

In the process of establishing an initial diagnosis of potential disease, a collection of relevant medical data can make a significant difference that enables doctors to make faster and more accurate decisions. For doctors, it is nowadays hard to consider a higher volume of data in the diagnosis procedure, or in determining the right treatment.

The diagnostics is a complicated process in which it is essential to have the right information available at the right time. For better decision making about patient diagnosis, we recommend an approach that compares existing patients with the new one and finds the most similar cases from history. This mechanism is known as Case-based reasoning (CBR) [1] which associates human reasoning with analogy. We implemented this method into an application that allows its users to reasonably compare new patients to the ones that already have their correct diagnosis. The use of CBR in the medical field is currently reviving. The knowledge base of medical knowledge is changing, and doctors have different approaches. The fact that the CBR methodology very much resembles the doctor's thinking suggests the successful use of CBR in the medical domain [7]. Cardiovascular diseases are still the leading cause of mortality not only in Slovak Republic, but in the world as well [10]. It was expected,

¹ Corresponding Author, Department of Cybernetic and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovakia; E-mail: ludmila.pusztova.2@tuke.sk.

when many innovative diagnostic and therapeutic approaches were introduced, that they will be accompanied by significant improvement of this situation. In contrary, risk profile of population worsened in the course of last years. Currently available modern diagnostic devices and methods can be used for early screening of patients, thereby speeding up patient selection for eventual intervention testing and indicating effective therapy. There is an unanswered question that properly indicated and evaluated new methods could reduce further investigations that unnecessarily prolong the diagnostic time for invasive-intervention strategies.

In our approach we adopted the CBR methodology (briefly described in the following section 2), utilizing k NN method for finding the most similar cases from history in the RETRIEVE phase, but also in an original way within the REVISE phase (more detailed description of the application design is provided in section 3). An important advantage of our application is visualization. There are more types of data visualizations, among them e.g. visualizations based on Principal Components Analysis (PCA). Some of the available visualizations as well as other implementation details are described in section 4.

2. Methodology

Case-based reasoning (CBR) is a problem-solving paradigm that is recognized and well-established method of medical science today [2]. To find a solution to a new problem, people try to use their own experience from dealing with previous tasks and adapt the old answers to fit the current situation [3]. Therefore, the main principle of the CBR is the assumption that solutions to similar problems should also be similar.

The CBR process consists of four main phases [4]:

- **RETRIEVE** – finding one or several of the most similar cases to the current problem,
- **REUSE/ADAPT** – reapplying the knowledge that was gained from the found cases to solve the actual problem,
- **REVISE** – checking the correctness of the proposed solution and modifying the solution if needed,
- **RETAIN** – saving the currently solved case for potential future use.

Algorithm k -Nearest Neighbors (k NN) was used to find the most similar previous cases. It is a versatile supervised machine learning algorithm that can be applied to solve problems leading to both classification and regression tasks. Fundamentally, algorithm k NN [5] assumes that cases close to each other, utilizing some suitable metric, are similar.

3. Application design

The goal of this application is to make the best use of anonymized collected medical data about real patients with cardiovascular diseases. It will serve as a decision support system for cardiologists. The main benefit of this application lies in helping the doctors to determine diagnosis faster and more precisely. Specifically, it should assist them

while deciding whether the current patient shows signs of potential cardiovascular disease and needs more exact examination such as coronary angiography.

The CBR cycle begins with the user entering medical data about a new patient who underwent various examinations. Entering the input data triggers the search for the most similar previous cases, which represents the first phase of the CBR cycle called **RETRIEVE**. This step utilizes the calculation of standard distance metrics such as Euclidean, Manhattan or Mahalanobis distance. We considered regarding the number of similar cases to be returned, the following options: only one case, an arbitrary number from interval $<1;10>$ or all the cases that are more similar than a given threshold. In the third scenario, there could be a situation when all of the known cases would be very different than the current one which means that none of them would be retrieved. Therefore, the combination of both options seems to be the most appropriate. At the same time, we also want to provide information about the calculated similarity of the cases.

The following phase, which is known as **REUSE/ADAPT**, uses the prevailing value of the target attribute as a solution for the new case. If the suggested value indicates that the new patient has a low risk of suffering from cardiovascular disease, the CBR cycle continues to the next phase.

The third phase is called **REVISE** because of the verification of the suggested diagnosis. We implemented the following approach. The application selects ten known cases that are the most similar to the current new case for each of the existing classes (i.e. for all different values of the target attribute). The calculated distances of the nearest cases from the current patient are then summed up for each of the classes of the target attribute and the class with the lowest result is marked as a revised suggested solution. If the newly proposed solution is the same as the initially suggested diagnosis (utilizing k NN), the user can consider this solution as a more trustworthy one.

We proposed also a different approach to the adaptation phase of the CBR cycle in [6]. Our approach there is based on the comparison of stored decision rules with the new one. Different parameters of the most relevant decision rule are identified and subsequently adapted to the new case.

4. Implementation

The proposed application is being implemented in the R language with the help of the Shiny package. The application makes use of the previously collected known cases which are stored in a relational table in a “.csv” file format.

The application currently consists of five different tabs. The user can navigate between them through the menu positioned at the top of the page. The first tab is called “The New Case” (see Fig. 1) and contains the main functionality of the application. The user has to enter necessary personal information, such as year of birth and gender of the new patient, to start the process of finding similar cases and suggesting the diagnosis. The number of cases taken into account can be any number in the interval $<1;10>$ with the default setting being the number 5. The information about the current patient and similar cases can be exported via the download button and saved in a “.docx” file format.

There are also five different options concerning the type of measure used for calculation of the distance between cases:

difference between the two PCA visualizations is that in the second visualization the user can alter the number of attributes that are selected for the calculation of the PCA components.



Figure 2. “PCA - Principal Component Analysis” tab.

5. Summary

In this paper we presented an original decision support system for cardiologists, which is based on the CBR methodology. k NN algorithm is implemented both, in the RETRIEVE as well as in the REVISE phases, in each of them in a different way. Moreover, CBR cycle is supported by a set of visualizations, which have been designed in tight cooperation with the target user, a cardiologist. Based on his feedback on the first version of our system, it seems to be very useful and promising. In future work we plan to evaluate our CBR system also in a user’s study with cardiologists.

Acknowledgments: This work was supported by the Slovak Research and Development Agency under grants no. APVV-16-0213 and APVV-17-0550.

References

- [1] M. Richter, R. Weber, *Case-Based Reasoning, A Textbook*, Springer, Heidelberg, 2013.
- [2] K. Bach, C. Marling, *Case-Based Reasoning Research and Development*, Springer, Heidelberg, 2019.
- [3] S. Begum, M. Ahmed, P. Funk, N. Yiong & M. Folke, Case-Based Reasoning Systems in the Health Sciences: A Survey of Recent Trends and Developments, *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews*, **41** (2011), 421–434.
- [4] N. Choudhury, S. Begum, A Survey on Case-based Reasoning in Medicine, *International Journal of Advanced Computer Science and Applications* **7** (2016), 136–144.
- [5] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Transact Info Theory*, **13** (1967), 21–27
- [6] L. Pusztová, F. Babič, J. Paralič & Z. Paraličová, How to Improve the Adaptation Phase of the CBR in the Medical Domain, *Lecture Notes in Computer Science* **11713** (2019), 168–177.
- [7] N. Choudhury & A. Shahin, A Survey on Case-based Reasoning in Medicine, *International Journal of Advanced Computer Science and Applications* **7** (2016), 136–144.
- [8] A. Salem, M. Roushdy & R. Hodhod, A Case Based Expert System for Supporting Diagnosis of Heart Diseases, *AIML Journal* **5** (2005), 33–39.
- [9] S. Wartner, M. Wiesinger-Widi, D. Girardi, D. Furthner & K. Schmitt K, Semi-automated Quality Assurance for Domain-Expert-Driven Data Exploration – An Application to Principal Component Analysis, *Lecture Notes in Computer Science* **11713** (2019), 128–146.
- [10] G. Backer, P. Jankovski, K. Kotseva, E. Mirrahimov, Z. Reiner, L. Ryden et al., Management of dyslipidemia in patients with coronary heart disease: Results from the ESC-EORP EUROASPIRE V survey in 27 countries, *Atherosclerosis* **285** (2019), 135–146

Clinical Decision Support Systems in Nursing Homes: A Scoping Review

Abir ABDELLATIF^{a,b,c,1}, Jacques BOUAUD^{d,a}, Duy NGHIEM^b, Carmelo LAFUENTE-LAFUENTE^{e,b}, Joël BELMIN^{e,b}, Brigitte SEROUSSI^{a,f}

^a Sorbonne Université, Université Paris 13, Sorbonne Paris Cité, INSERM, UMR S_1142, LIMICS, Paris, France

^b AP-HP, Hôpital Charles-Foix, Ivry-sur-Seine, France

^c Teranga Software, Paris, France

^d AP-HP, DRCI, Paris, France

^e Sorbonne Université, Paris, France

^f AP-HP, Hôpital Tenon, Paris, France

Abstract. The world population is dramatically ageing, resulting in an increase of the prevalence of older dependent adults living in nursing homes (NHs). Because of insufficient resources in NHs, and nurses' lack of time and knowledge, adverse events, most of them being preventable, are often reported. Clinical decision support systems (CDSSs) have proven to improve the quality of care in various healthcare settings such as hospitals and primary care centers. However, the use of CDSSs in NHs is still limited and little is known about their influence on nursing practices and NH residents' clinical outcomes. We conducted a scoping review of the literature to evaluate CDSS use in NHs. Out of 1,231 retrieved papers, 15 studies were retrieved which assess 10 CDSSs applied to pressure ulcers and malnutrition prevention, drug prescription, and disease management. This review showed CDSSs could be effective in NHs for improving routine clinical practice and patient outcomes, but research is still needed to implement effective CDSSs in NHs.

Keywords. Clinical decision support systems, Nursing homes, Geriatrics, Clinical Outcomes, Evaluation, Quality of care

1. Introduction

Because of the ageing population and the rise in the prevalence of dependent older adults, an increasing number of aged people currently live in nursing homes (NHs). Studies have reported that NH residents are at risk of developing adverse events, many are preventable (e.g. pressure ulcers, urinary tract infections, falls, adverse drug events) [1]. This is often explained by insufficient staff and resources in NHs, and nurses' lack of time and knowledge [2].

Electronic Health Record systems (EHRs) contain patient's data, including nursing records and care track. EHRs have proven to provide accurate, up-to-date, and

¹ Corresponding Author: Abir Abdellatif, Hôpital Charles-Foix, 7 avenue de la République, 94200 Ivry-sur-Seine, France. E-mail : abir.abdellatif@gmail.com

holistic information about patients at the point of care, enabling more coordinated and efficient care. The use of clinical decision support systems (CDSSs) connected to EHRs have demonstrated to improve the quality of care in both hospitals and primary care settings [3]. However, the use of CDSSs in NHs is still limited and little is known about their influence on nursing practices and NH residents clinical outcomes [4].

Teranga² is a French vendor of an EHR called “NETSoins” used in about 40% of NHs nationwide. One objective is to develop a CDSS connected with NETSoins to improve the knowledge of nurses and support them in providing better care to NH residents. Prior to this development, we conducted a scoping review of the literature to analyze CDSSs used in NHs and assess the key factors of these systems regarding nurse satisfaction, care process improvement, and NH residents clinical outcomes.

2. Methods

PubMed was searched between 2010 and 2019 for studies describing CDSSs used in NHs. We used a combination of MeSH terms and key words to account for the management of elderly in NHs (i.e., “Geriatrics”, “Long-Term Care”, “Nursing homes”, “Residential Facilities”, “Care homes”) and the use of CDSSs (i.e., “Decision support techniques”, “Decision support system, Management”, “Decision support system, Clinical”).

Two of the authors (AA, DN) independently selected original articles describing and evaluating the implementation and use of CDSSs in NHs on the basis of titles and abstracts. The reference list of selected papers was also examined.

Selected articles were analyzed according to (i) the specific clinical application of the CDSS, (ii) the design of the study (type, CDSS users, sample size), (iii) CDSS features (reasoning process and users’ interfaces), (iv) impact of the CDSS, and (v) CDSS users’ acceptance and adoption in routine practice beyond the intervention period. We included observational studies (OSs), randomized (RCTs), and non-randomized controlled trials (NRCTs).

3. Results

The query used to search PubMed retrieved a total of 1,231 articles leading to 26 potentially relevant publications for full-text review after the screening step. Eleven publications were excluded (CDSS was applied to primary care, no computerized CDSS was used, no outcome was assessed) leading to 15 articles [4–18] that met our criteria for review.

The 15 selected studies assessed 10 different CDSSs that were gathered for the analysis according to their clinical application: CDSSs for prevention, CDSSs for drug prescription, and CDSSs for disease management. In all the retrieved studies, users’ adoption of CDSSs in routine practice beyond the intervention study period was not well described and was finally not analyzed in this paper.

² <https://www.teranga-software.com>

3.1. CDSSs for pressure ulcers and malnutrition prevention

We retrieved five studies [4–8] assessing the implementation of three CDSSs to support pressure ulcers and malnutrition prevention. Two studies were OSs [5,8], two were RCTs [4,7], and one was a NRCT [6]. All studies enrolled nurses and nursing aides as primary CDSS users. All CDSSs were guideline-based and used the display of alerts automatically triggered when risk assessment instruments (e.g. the Risk Assessment Pressure Scale for pressure ulcer risk screening or the Mini Nutritional Assessment scale for screening nutritional status) exceeded predefined thresholds. One CDSS [4–6] is integrated into the NH EHR, the two others [7,8] are standalone systems that need to be activated by users. One CDSS [8] is a patient monitoring system to facilitate nursing staff implementation of standard care for pressure ulcer/injury prevention. The system operates from data on frequency and position of sensors worn on each resident's anterior chest, to estimate the nursing staff compliance with repositioning standard of care, before and after visual monitors were activated to cue staff.

Globally, using the CDSS lowered significantly the prevalence of malnourished residents (reduction from 28.8 to 19.8%) [4], and pressure ulcers incidence (reduction from 15.1% to 7.1%) [7]. The three CDSSs were associated with an improvement of care processes (better compliance of nursing care with clinical practice guidelines, especially concerning NH residents repositioning, improvement of practitioners' attitude and nursing culture) [5–8]. User reviews were mixed, diversely reporting ease of use, lack of training, and resistance to computer use.

3.2. CDSSs for drug prescription: medication review and daily drug prescription

We made the difference between CDSSs supporting daily drug prescription and those for “medication reviews”. A medication review is a structured analysis and re-consideration of the entire patient's drug treatment aiming at simplifying and optimizing it, usually performed at least once a year in geriatrics by both physician and pharmacist.

Five studies [11,12,16–18] (among which four OSs and one NRCT), evaluated two CDSSs for medication review. Two studies [11,12] enrolled pharmacists as primary users of the CDSS whereas nurses were the CDSS principal users in the other studies [16–18]. Both CDSSs were guideline-based. One [11,12] is a standalone system and operates to support pharmacist review of physician orders. In case of drug mis-prescription, alerts are automatically triggered and displayed to the pharmacist who contacts the prescriber and indicates how to revise the wrong prescription. The second CDSS [16–18] is connected to the NH EHR. It criticizes drug prescriptions and alerts nurses about drug-related problems. All studies measured the effects of using the CDSS on care processes such as prescriptions and observed clinical practices compliance with guidelines, as well as interpersonal collaboration. The use of the CDSS was beneficial in all studies. Users were mostly satisfied by the CDSS except in one study [12], where they reported alert fatigue.

Three studies [13–15] (two OSs and one RCT) evaluated the implementation of three CDSSs to support daily drug prescription. The primary users of CDSSs were pharmacists in two studies [14,15], and physicians in the third study [13]. The knowledge bases of these systems are elaborated from clinical practice guidelines and expert opinion. All CDSSs display critics attached to non-compliant drug dosing or

prescription. One of the three CDSSs [9] is connected with the NH EHR and generates informative dosing messages or alerts for prescribers. The two other CDSSs are standalone systems where patient data have to be entered to obtain decision support. Warning messages are displayed to criticize physicians' prescriptions. CDSSs work with emails or phone calls between physicians and pharmacists.

CDSSs were successful in all studies to improve processes (e.g. increase of drug prescription compliance) and organizational aspects (e.g., time saving, team working). No clinical outcome was measured in any of the studies. Users of one study [13] expressed their dissatisfaction about unnecessary alerts (displayed in 50% of the cases).

3.3. CDSSs for disease management

Two studies (both were OSs) [9,10] assessed two CDSSs for disease management. The first CDSS [9] is applied to help data tracking and analysis to be used by physicians (tabular view for medications, weight-trending graph, and symptom analysis). The second CDSS [10] is applied to the management of urinary tract infections and is developed following Goal-Directed Design principles to support NH staff in differentiating urinary tract infections from asymptomatic bacteriuria, thus reducing antibiotic days.

The first CDSS [9] is connected to the NH EHR. It relies on a guideline-based knowledge base. The second CDSS [10] is a web-based application. Four information needs were identified from focus groups (guidance regarding resident assessment, communication with providers, care planning, and urine culture interpretation) and encoded in the knowledge base of the system (also including clinical practice guidelines). Both CDSSs provided educational recommendations. Both CDSS were considered useful for symptom tracking, visual information, facility of data entry, and ease of use.

4. Discussion and conclusion

Health information technology tools such as EHRs and CDSSs have proven to improve the quality of care in hospital settings. However, our literature review showed that these tools are not yet widely used in NHs despite a great number of residents and insufficient staff and resources, e.g., lack of doctors. On the last decade, only 15 papers describing the NH use of 10 different CDSSs have been retrieved. Most of the CDSSs were stand alone systems (six out of 10), this lack of interoperability being reported as being an issue in previous studies. Systems were poorly described in the retrieved papers and it was difficult to understand how they were developed (data processing, user's interfaces, etc.). For instance, we could not assess whether factors such as requiring practitioners to provide reasons when over-riding or providing advice concurrently to patients and practitioners, known to be winning factors to improve guideline compliance and care quality [3], were implemented. Only a few studies were RCTs (3 out of 15), evidencing positive clinical outcomes (significant reduction in the prevalence of malnourished residents and pressure ulcers incidence), and most of CDSSs were judged as satisfactory (ease of use, time saving, team working) even if some critics were also reported (alert fatigue). The number of health conditions addressed is currently quite limited although there is, undoubtedly, a great potential to expand CDSS applications, for instance to the management of osteoporosis or to fall

prevention. However, since elderly are more likely to experience adverse drug reactions, many studies were focused on drug prescription improvement.

Whether they are linked or not to an EHR, used by physicians, nurses, or pharmacists, generating alerts at the point of care or not, this review suggests CDSSs can be effective in NHs to improve nurses' daily practice and residents' clinical outcomes in several domains. However, more research is still needed for implementing effective CDSSs in NHs.

References

- [1] Thomas EJ, Brennan TA. Incidence and types of preventable adverse events in elderly patients: population based review of medical records. *BMJ* **320** (2000), 741–744.
- [2] White EM, Aiken LH, McHugh MD. Registered Nurse Burnout, Job Dissatisfaction, and Missed Care in Nursing Homes. *J Am Geriatr Soc* **67** (2019), 2065–2071.
- [3] Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. *BMJ* **346** (2013), f657–f657.
- [4] Fossum M, Alexander GL, Ehnfors M, Ehrenberg A. Effects of a computerized decision support system on pressure ulcers and malnutrition in nursing homes for the elderly. *Int J Med Inf* **80** (2011), 607–617.
- [5] Fossum M, Ehnfors M, Fruhling A, Ehrenberg A. An Evaluation of the Usability of a Computerized Decision Support System for Nursing Homes. *Appl Clin Inform* **02** (2011), 420–436.
- [6] Fossum M, Ehnfors M, Svensson E, Hansen LM, Ehrenberg A. Effects of a computerized decision support system on care planning for pressure ulcers and malnutrition in nursing homes: An intervention study. *Int J Med Inform* **82/10** (2013), 911–921.
- [7] Beeckman D, Clays E, Van Hecke A, Vanderwee K, et al. A multi-faceted tailored strategy to implement an electronic clinical decision support system for pressure ulcer prevention in nursing homes: a two-armed randomized controlled trial. *Int J Nurs Stud* **50** (2013), 475–486.
- [8] Yap TL, Kennerly SM, Ly K. Pressure Injury Prevention: Outcomes and Challenges to Use of Resident Monitoring Technology in a Nursing Home. *J Wound Ostomy Continence Nurs* **46** (2019), 207–213.
- [9] Zhu W, Luo L, Jain T, Boxer RS, Cui L, Zhang G-Q. DCDS: A Real-time Data Capture and Personalized Decision Support System for Heart Failure Patients in Skilled Nursing Facilities *AMIA Annu Symp Proc.* (2016), 2100–2109.
- [10] Jones W, Drake C, Mack D, Reeder B, Trautner B, Wald HL. Developing Mobile Clinical Decision Support for Nursing Home Staff Assessment of Urinary Tract Infection using Goal-Directed Design. *Appl Clin Inform* **8** (2017), 632–650.
- [11] de Wit HAJM, Mestres Gonzalvo C, Hurkens KPGM, Mulder WJ, Janknegt R, Verhey FR, et al. Development of a computer system to support medication reviews in nursing homes. *Int J Clin Pharm* **35** (2013), 668–672.
- [12] de Wit HAJM, Mestres Gonzalvo C, Cardenas J, Derijks HJ, Janknegt R, van der Kuy P-HM, et al. Evaluation of clinical rules in a standalone pharmacy based clinical decision support system for hospitalized and nursing home patients. *Int J Med Inf* **84** (2015), 396–405.
- [13] Donovan JL, Kanaan AO, Thomson MS, Rochon P, Lee M, Gavendo L, et al. Effect of clinical decision support on psychotropic medication prescribing in the long-term care setting. *J Am Geriatr Soc* **58** (2010), 1005–1007.
- [14] Kennedy CC, Campbell G, Garg AX, Dolovich L, Stroud JB, McCallum RE, et al. Piloting a renal drug alert system for prescribing to residents in long-term care. *J Am Geriatr Soc* **59** (2011), 1757–1759.
- [15] Ulfvarson J, Bastholm Rahmner P, Fastbom J, Sjövikar S, Andersén Karlsson E. Medication reviews with computerised expert support: Evaluation of a method to improve the quality of drug utilisation in the elderly. *Int J Health Care Qual Assur* **23** (2010), 571–582.
- [16] Johansson-Pajala R-M, Gustafsson L-K, Jorsäter Blomgren K, Fastbom J, Martin L. Nurses' use of computerised decision support systems affects drug monitoring in nursing homes. *J Nurs Manag* **25** (2017), 56–64.
- [17] Johansson-Pajala R-M, Martin L, Jorsäter Blomgren K. Registered nurses' use of computerised decision support in medication reviews. *Int J Health Care Qual Assur* **31** (2018), 531–544.
- [18] Johansson-Pajala R-M. Conditions for the Successful Implementation of Computer-Aided Drug Monitoring From Registered Nurses' Perspective-A Case Site Analysis. *Comput Inform Nurs CIN* **37** (2019), 196 - 202.

Comparison of Unplanned 30-Day Readmission Prediction Models, Based on Hospital Warehouse and Demographic Data

Thibault DHALLUIN ^{a,1}, Aurélie BANNAY ^b, Pierre LEMORDANT ^a, Emmanuelle SYLVESTRE ^a, Emmanuel CHAZARD ^c, Marc CUGGIA ^a and Guillaume BOUZILLE ^a

^aUniv Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France

^bCHRU-Nancy, Université de Lorraine, Service d'évaluation et d'information médicales, F54000, France

^cUniv Lille, CHU Lille, CERIMEA2694, F-59000 Lille, France

Abstract. Anticipating unplanned hospital readmission episodes is a safety and medico-economic issue. We compared statistics (Logistic Regression) and machine learning algorithms (Gradient Boosting, Random Forest, and Neural Network) for predicting the risk of all-cause, 30-day hospital readmission using data from the clinical data warehouse of Rennes and from other sources. The dataset included hospital stays based on the criteria of the French national methodology for the 30-day readmission rate (i.e., patients older than 18 years, geolocation, no iterative stays, and no hospitalization for palliative care), with a similar pre-processing for all algorithms. We calculated the area under the ROC curve (AUC) for 30-day readmission prediction by each model. In total, we included 259114 hospital stays, with a readmission rate of 8.8%. The AUC was 0.61 for the Logistic Regression, 0.69 for the Gradient Boosting, 0.69 for the Random Forest, and 0.62 for the Neural Network model. We obtained the best performance and reproducibility to predict readmissions with Random Forest, and found that the algorithms performed better when data came from different sources.

Keywords. Medical Informatics, Data Warehousing, Supervised Machine Learning, Patient Readmission/statistics and numerical data

1. Introduction

Controlling the proportion of unexpected readmissions is important for patient comfort and safety and represents a medico-economic issue. In 2009, the 30-day readmission rate in the United States was about 20% (13.3% to 23.2%, depending on the state [1]), and the cost of these readmissions was estimated at 17.4 billion dollars. In France, the readmission rate varies between 12% and 14%, according to the methodology [2].

To comprehensively address this problem in the context of population ageing and the increasing prevalence of chronic diseases, it is important to take into account the

¹ Corresponding Author, Thibault DHALLUIN, E-mail : Thibault.DHALLUIN@chu-rennes.fr

care pathway between hospital and primary care. Indeed, the days following the hospital discharge are at risk of adverse events [3–4], and therefore, a better communication between caregivers is essential. A systematic review highlighted that some 30-day readmission episodes are avoidable, but their proportion greatly varies partly in function of the (objective/subjective) criteria used to define such avoidability [5]. Nevertheless, the readmission rate is widely used as indicator to evaluate possible improvements in medical care.

Public health policies have been put in place to control the 30-day readmission rates in the United States [6], France, Germany and Denmark [7]. Consequently, research has been focused on developing methods to anticipate unplanned readmissions, by identifying relevant clinical indicators, and by developing predictive models to integrate these indicators.

The adoption of electronic health records (including laboratory data) represents an unprecedented source of data that must be exploited to better understand the patient care pathway, both descriptively and analytically. The emergence of automatic learning methods allows using these data to anticipate the patient trajectories. The most widespread predictive models [8–10] are based on classical statistical approaches that are widely used in medical research, such as logistic regression analysis. More recently, machine learning methods have started to be tested (e.g., Gradient Boosting, Random Forest, Neural Networks). Therefore, the contribution of each of these methods must be now assessed in similar conditions.

The goal of this study was to compare different approaches to predict the 30-day readmission risk of hospitalized patients using massive hospital and geo-demographic data.

2. Methods

The included hospital stays had an admission date that ranged from January 1, 2013 to December 31, 2018. This dataset was then partitioned into a training set (admission date before December 31, 2016) and a test set (admission date after January 1, 2017). To ensure the results comparability with the French national indicator for 30-day readmission, the selection criteria were those of the ATIH RH30 national methodology (Agence Technique de l'Information sur l'Hospitalisation - Les réhospitalisations à 30 jours) [11], except for the geolocation data that were used at the place of the geographical code. Specifically, the patients included were aged 18 years or over and received obstetrical, surgical or medical care. Patients with no geolocation in mainland France and patients in palliative care settings were excluded. Hospital stays with a different mode of entry than from home and iterative stays (chemotherapy session, radiotherapy, transplant context, renal dialysis sessions, cataract surgery) also were excluded. A readmission was defined as an unplanned hospital stay within 30 days after the end of the previous stay.

Data were extracted from the Rennes clinical data warehouse [12] that contains most of the computerized medical records, including clinical notes, drug prescriptions, laboratory test records and administrative data. All information about a patient's hospital stay was de-identified and organized in structured and unstructured data in the electronic medical record. If the patient had several hospital stays, all hospital stays were combined in a single electronic medical record. Diagnoses were coded using the French version of the International Classification of Diseases, 10th Edition (ICD-10)

that incorporates the specific French adjustments. Medical and surgical procedures were coded according to the French classification of clinical procedures (CCAM classification) [13]. Drug prescriptions were coded using the Anatomical, Therapeutic and Chemical (ATC) code. The potentially significant covariates extracted from the clinical data warehouse were: age, sex, length of stay, number of previous hospitalizations, illness severity in the French version of diagnosis-related group (DRG) classification, major diagnostic categories, medical diagnosis and comorbidities (ICD-10), medical and/or surgical procedures (CCAM), hospital department and laboratory data (still organized according to a local thesaurus).

Demographic data, aggregated at the municipal and sub-municipal level [14], were merged with the patient using the geometric map background and the corresponding geolocation: median pre-tax household income during the year, part of ≥ 15 -year-old population, education level, unemployment rate for the 15 to 64-year-old population, and socio-professional categories for the 15 to 64-year-old population.

Data handling and pre-processing were performed on R with 'tidyverse' (version 1.2.1), 'data.table' (version 1.12.2) and 'recipes' (version 0.1.6). The compared algorithms were the most frequently found in the literature: logistic regression ('stats' version 3.5.0), random forest ('randomForest' version 4.6-14), gradient boosting ('gbm' version 6.1.1), and neural network ('nnet' version 7.3-12).

The main outcome was the area under the ROC curve (AUC) and the secondary outcomes were sensitivity, specificity, positive and negative predictive value of cut-point closest to the top-left corner of ROC space. As the models did not yet have a specific use, there was no reason to focus on sensitivity (to screen for stays at risk of readmission), and specificity (to target specifically stays with readmission).

The importance of the different covariates (clinical/geo-demographic data) was assessed according to criteria adapted to each algorithm: odds ratio for logistic regression, relative influence for gradient boosting, Gini index for random forest, and Garson algorithm for neural network [15].

3. Results

Among the 410 837 stays available in the clinical data warehouse, 259114 were retained (193913 for the training set and 65201 for the test set) based on the inclusion and non-inclusion criteria. The main non-inclusion criteria were patients younger than 18 years (76670 stays excluded), absence of geolocation (18454 stays excluded), iterative stays (14477 stays excluded), and palliative care (5750 stays excluded). The 30-day readmission rate was 8.8%. Readmitted patients were on average 4 years older than non-readmitted patients, and with a higher proportion of males (58.3% versus 49.2%). Demographic data were comparable in readmitted and non-readmitted patients.

The AUC was 0.61 for logistic regression, 0.69 for gradient boosting, 0.69 for random forest, and 0.62 for neural networks (Figure 1). Logistic regression and gradient boosting placed the most importance on post-emergency units, number of iterative hospitalizations, disease severity (DRG classification) and co-morbidities. The regression model became robust by adding a stepwise cross-validation. Random forest focused on other variables, including demographic data. In addition, random forest gave the most robust results, regardless of data sharing, with constant performance and use of covariates.

Table 1. Main characteristics of the patients of the included stays.

Readmission	YES	NO
N	22893	236221
Age (mean (sd))	60.30 (19.34)	56.46 (20.67)
Sex, male (%)	13355 (58.3)	116329 (49.2)
Length of stay (mean (sd))	6.18 (10.47)	4.81 (8.14)
Number of previous stays (mean (sd))	2.58 (3.27)	1.82 (1.85)
Median pre-tax household income (mean (sd))	20654.29 (3274.12)	20643.86 (3210.86)
Higher degree rate (mean (sd))	0.30 (0.13)	0.30 (0.13)
Unemployment rate (mean (sd))	0.12 (0.06)	0.12 (0.06)

Table 2. Prediction performances.

Algorithm	TP	TN	FP	FN	Se	Sp	PPV	NPV	AUC
Logistic Regression	3011	37187	22325	2678	0.53	0.62	0.12	0.93	0.61
Gradient Boosting	3531	39065	20447	2158	0.62	0.66	0.15	0.95	0.69
Random Forest	3528	39075	20437	2161	0.62	0.66	0.15	0.95	0.69
Neural Network	3467	33911	25601	2222	0.61	0.57	0.12	0.94	0.62

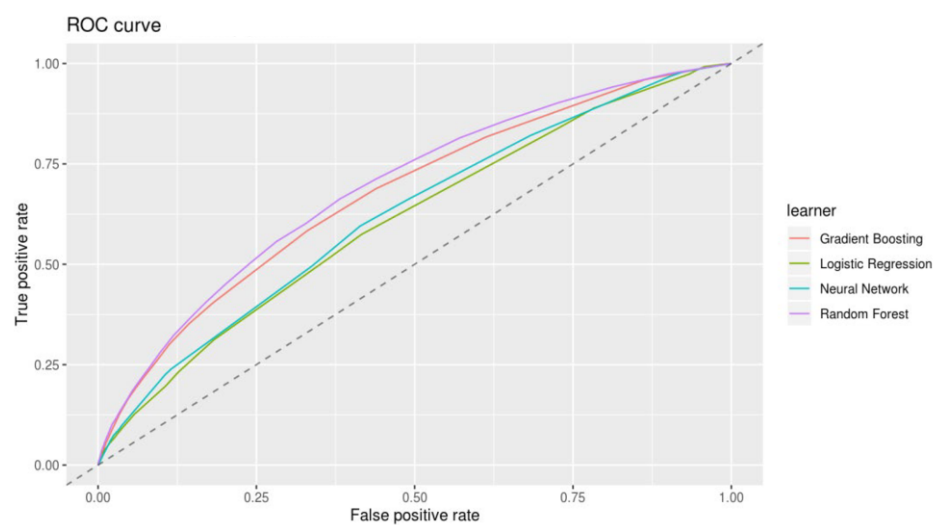


Figure 1. ROC curve.

4. Discussion

This is the first comparison of 30-day predictive readmission models in France using clinical data warehouse and geo-demographic data. In addition, this study used a methodology as similar as possible to the definition of the French national 30-day readmission indicator, while taking advantage of data from different sources.

There is a wide literature on the subject of 30-day hospital readmission prediction models. The AUC values of our prediction models are comparable to the 0.71 found in the literature [8]. The community state of health, healthcare delivery/access, hospital management, computerization of departments, data integration, and advancement of the data warehouse are among the many factors that affect the results of the obtained models. The inclusion criteria and the choice to focus on all-cause readmissions also

influence the prediction model performance. Indeed, the algorithm used is an end-of-chain parameter that depends very strongly on the learning data quality.

5. Conclusion

In conclusion, the predictive models displayed comparable performances (AUC from 0.61 to 0.69). Random forest showed the best and most reproducible performance compared with the other models. The importance of the included variables differs among models and merging external data can improve their performance.

References

- [1] S.F. Jencks, M.V. Williams, and E.A. Coleman, Rehospitalizations among Patients in the Medicare Fee-for-Service Program, *New England Journal of Medicine*. **360** (2009) 1418–1428. doi:10.1056/NEJMs0803563.
- [2] M. Gusmano, V. Rodwin, D. Weisz, C. Jonathan, and C. Quantin, A Comparative Analysis of Hospital Readmissions in France and the US, *Journal of Comparative Policy Analysis: Research and Practice*. (2015) 1–15. doi:10.1080/13876988.2015.1058547.
- [3] A.J. Forster, H.J. Murff, J.F. Peterson, T.K. Gandhi, and D.W. Bates, The incidence and severity of adverse events affecting patients after discharge from the hospital, *Ann. Intern. Med.* **138** (2003) 161–167. doi:10.7326/0003-4819-138-3-200302040-00007.
- [4] J.L. Schnipper, J.L. Kirwin, M.C. Cotugno, S.A. Wahlstrom, B.A. Brown, E. Tarvin, A. Kachalia, M. Horng, C.L. Roy, S.C. McKean, and D.W. Bates, Role of Pharmacist Counseling in Preventing Adverse Drug Events After Hospitalization, *Arch Intern Med.* **166** (2006) 565–571. doi:10.1001/archinte.166.5.565.
- [5] C. van Walraven, C. Bennett, A. Jennings, P.C. Austin, and A.J. Forster, Proportion of hospital readmissions deemed avoidable: a systematic review, *CMAJ*. **183** (2011) E391–E402. doi:10.1503/cmaj.101860.
- [6] C. for Medicare, M.S. 7500 S.B. Baltimore, and M. Usa, Readmissions-Reduction-Program, (2019). <https://www.cms.gov/medicare/medicare-fee-for-service-payment/acuteinpatientpps/readmissions-reduction-program.html> (accessed July 24, 2019).
- [7] S.R. Kristensen, M. Bech, and W. Quentin, A roadmap for comparing readmission policies with application to Denmark, England, Germany and the United States, *Health Policy*. **119** (2015) 264–273. doi:10.1016/j.healthpol.2014.12.009.
- [8] A. Artetxe, A. Beristain, and M. Graña, Predictive models for hospital readmission risk: A systematic review of methods, *Computer Methods and Programs in Biomedicine*. **164** (2018) 49–64. doi:10.1016/j.cmpb.2018.06.006.
- [9] H. Zhou, P.R. Della, P. Roberts, L. Goh, and S.S. Dhaliwal, Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review, *BMJ Open*. **6** (2016). doi:10.1136/bmjopen-2016-011060.
- [10] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, and S. Kripalani, Risk prediction models for hospital readmission: A systematic review, *JAMA*. **306** (2011) 1688–1698. doi:10.1001/jama.2011.1515.
- [11] DGOS, Les indicateurs de réhospitalisation et de coordination, *Ministère des Solidarités et de la Santé*. (2019). <https://solidarites-sante.gouv.fr/soins-et-maladies/qualite-des-soins-et-pratiques/qualite/les-indicateurs/article/re-hospitalisation-coordination> (accessed September 3, 2019).
- [12] G. Bouzillé, E. Sylvestre, B. Campillo-Gimenez, E. Renault, T. Ledieu, D. Delamarre, and M. Cuggia, An Integrated Workflow For Secondary Use of Patient Data for Clinical Research., *Stud Health Technol Inform.* **216** (2015) 913–913.
- [13] CCAM en ligne - Consultation par chapitre, (n.d.). <https://www.ameli.fr/accueil-de-la-ccam/trouver-un-acte/consultation-par-chapitre.php> (accessed September 3, 2019).
- [14] Contours...Iris® | IGN - Espace professionnel, (n.d.). <http://www.professionnels.ign.fr/contoursiris> (accessed September 3, 2019).
- [15] G.D. Garson, Interpreting Neural-network Connection Weights, *AI Expert*. **6** (1991) 46–51.

Cross-Border Exchange of Clinical Data Using Archetype Concepts Compatible with the International Patient Summary

Dimitar TCHARAKTCHIEV^a, Evgeniy KRASTEV^{bl}, Patrick PETROSSIANS^c,
Simeon ABANOS^b, Hristo KYURKCHIEV^b, Petko KOVATCHEV^b

^aDepartment of Medical Informatics, Medical University, Sofia, Bulgaria

^bFaculty of Mathematics and Informatics, Sofia University St. Kliment Ohridsky, Bulgaria

^cCentre Hospitalier Universitaire de Liège, Liège, Belgium

Abstract. This paper proposes an approach and demonstrates its application for cross-border exchange of clinical documents oriented towards the use of archetype concepts and international patient summary standards adopted in the European Union. A novelty in this approach is the management of native XML instances of an archetype concept in the CEN 13606 standard by means of a native XML database and XML technologies. The computer experiments demonstrate that it is suitable for representing relatively small clinical datasets such as those describing rare diseases like the Acromegaly illness, where the semantic context in the relatively small number of symptoms is practicable to tag in terms of SNOMED-CT terminology codes. Additionally, we demonstrate that the semantically enriched information model can facilitate secondary use of clinical data by visualizing the execution of queries based on standard terminologies. Finally, the compatibility of the information model with the IPS standard enables sharing of clinical data among different information models.

Keywords. international patient summary, semantic interoperability, CEN 13606 archetype, e-health, rare disease register, native XML database, terminology

1. Introduction

European citizens enjoy a variety of rights when it comes to intra-EU mobility, including the right to seek care in another Member State. Health systems in the European Union aim to provide high-quality, cost-effective care. Directive 2011/24/EU on patients' rights in cross-border healthcare describes patients' right to receive health treatment, including planned procedures, in another Member State (MS) and to then get reimbursed domestically up to the level that would have been covered by their home State [1].

The realization of the full potential of the Directive's provisions is still in development. In March 2017, the Commission launched the European Reference Networks (ERNs), creating cross-border virtual networks of healthcare providers across the EU to tackle rare diseases and conditions [2]. The ERNs are supported by Clinical Patient Management System (CPMS) in providing knowledge and resources for

¹ Corresponding Author: email: dimitardt@gmail.com

improving the diagnosis and treatment of rare or low prevalence complex diseases [3]. CPMS establishes a common ground for collaboration between healthcare professionals in this subject area allowing them to upload patient cases in Rare Disease (RD) registers and share this way diagnosis, care and treatment procedures. However, only a Digital Imaging and Communications in Medicine (DICOM) Viewer for image consultations and a couple of IHE profiles for connection with Hospital Information Systems are provided, no other European or International standards for data exchange are implemented. A Central Metadata Repository (ERDRI.mdr) is established in 2019 in order to ensure semantic interoperability between existing RD registries [4]. It stores all data elements (metadata) used by the participating registries, including the names of the data elements (designations) and their definitions. Authorized users can create groups of hierarchical structures of new data elements in a limited number of languages (English, German, Spanish and French). This good principle of containerization is not related to any standard approved in the European Union like the CEN 13606 standard, where the reference model is composed of hierarchical structures. Moreover, the content of data elements may be represented with minor exceptions in an arbitrary way without a reference to external ontologies, terminologies or standards. This complicates substantially the exchange of data between European databases and national RD registers.

The situation requires prompt and adequate solutions really helping the clinicians to exchange clinical information without wasting time to translate and transcribe the already registered medical records. In our point of view the clinical data sets for several rare diseases that were accumulated in the last decades as part of existing national registries are suitable to start the data exchange, maintaining the clinical context. These datasets are structured clinical records and include several important variables [5]. Their exchange using European health informatics standards between existing European data bases and National RD registries can foster the way for implementation of semantic interoperability.

The objective of this study is to prove that the CEN 13606 Standard for Electronic Record Communication based on the Archetype paradigm and the International Patient Summary (IPS) together with the appropriate international nomenclatures and classifications successfully could provide a high degree of semantic interoperability. In the following section we present the methods we employ to achieve this objective. The results from computer experiments are discussed in Section 3 and the conclusions are outlined in Section 4.

2. Methods

In this paper we present an approach and demonstrate its application for cross-border exchange of pseudonymized clinical records between the Bulgarian Acromegaly register and the Liege acromegaly survey (LAS) database that aggregates the data from 14 European RD registries [6]. This approach is oriented towards the use of archetype CEN 13606 concepts and maintenance of compatibility with the International Patient Summary (IPS) described in the preliminary standard (FprEN 17269 [7]) and the technical specification (FprCEN/TS 17288 [8]).

International nomenclatures like SNOMED CT, LOINC, ICD 10 and ATC are used in the CEN 13606 Archetypes. A mapping is performed between the clinical concepts existing in the national databases and the respective codes in these nomenclatures for laboratory tests, procedures, diagnoses and medications. This allows to design an

archetype of the RD dataset compatible with the IPS standard enriched with semantic context (Fig. 1-left side) and describe it using the Archetype Description Language in CEN 13606. Unlike existing approaches that employ archetypes with relational databases we obtain a native XML instance of the obtained archetype(Fig. 1-right side). We refer to this XML instance as native because it is valid with respect to the CEN 13606 Reference model [9]. It allows us to represent RD datasets in terms of such instances with native XML database like existDB [10] and execute queries on the such archetype instances with XML technologies. Note, that an archetype query language is not provided as part of the CEN 13606 standard.

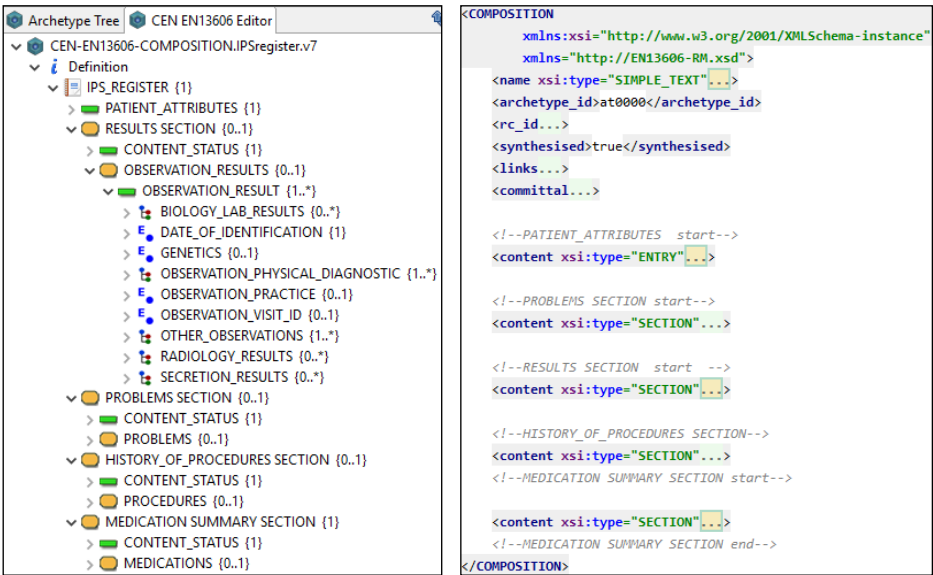


Figure 1. CEN 13606 Archetype of IPS sections(left) and its corresponding native XML instance (right).

3. Results

The methods described in the previous section have been applied to develop a client server web application with the NoSQL database existDB [10]. A dataset with real clinical data from the Bulgarian Acromegaly register is loaded as a “collection” of XML documents in this database. Each one of these documents is a native XML instance of the archetype shown in Fig.1. The client makes use of the web graphical interface on Fig.2 to create new native XML instances of the archetype and add data content in these instances. For example, Fig. 2 shows the recorded clinical data for symptoms observed during the Acromegaly therapy of a given Patient, where the semantic context is represented by SNOMED-CT codes [11].

The web application makes use of RESTful web services allowing the client to manage the collection of XML documents on existDB. It enables cross- border exchange of RD datasets for example, with the Liege acromegaly survey (LAS) database. Once the XML document is complete it can be uploaded in the LAS database. Moreover, this standard- based approach allows to query the RD dataset and visualize the results. For example, Fig. 3 shows visualization from two of the major Acromegaly symptoms during

the Acromegaly therapy of the selected Patient. This visualization allows us to conclude that therapy has been successful, and the Patient has fully recovered.

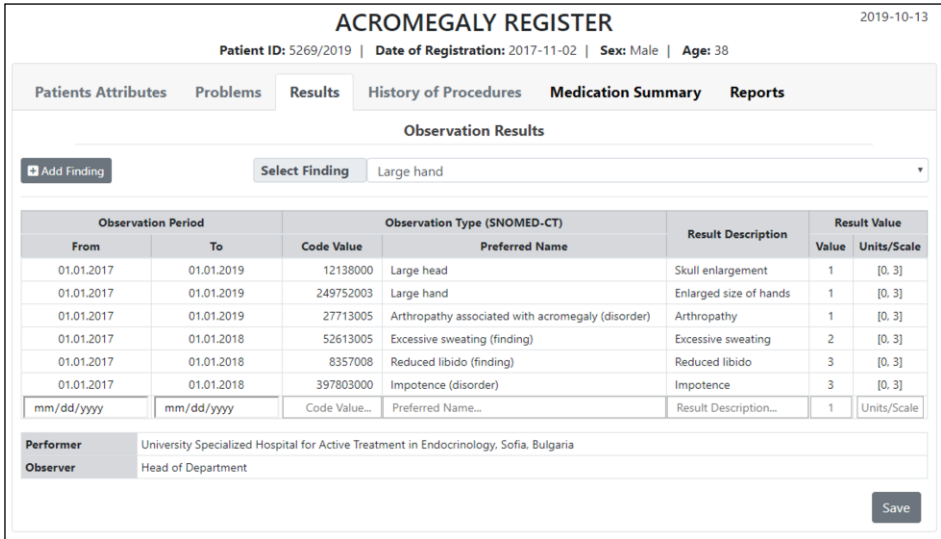


Figure 2. Web interface for management of XML instances of the IPS archetype.

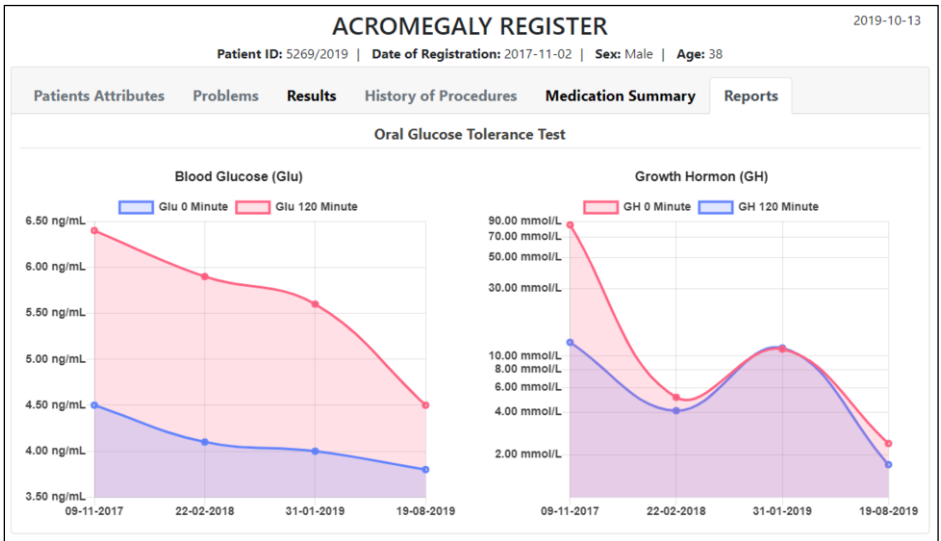


Figure 3. Visualization of selected clinical data for Acromegaly Symptoms.

4. Conclusion

In this paper we have proposed an approach and discussed the results of its software implementation for cross- border exchange of clinical data for rare diseases like the Acromegaly illness. The implementation of the proposed approach employs EN 13606

archetypes to build a model of a clinical document compatible with the recently proposed IPS standard FprEN 17269. The novelty in this approach is the creation of a native XML instance of a CEN 13606 archetype concept representing a clinical document model in terms of the Archetype Description Language. It allows to achieve semantic interoperability in exchange of RD data sets compatible with the IPS standard by means of native XML databases and XML technologies. The computer experiments show that this approach and its implementation are quite suitable for representing relatively small clinical datasets such as those describing rare diseases. For example, the semantic context in the limited number of symptoms in the Acromegaly Register is practicable to tag with SNOMED-CT terminology codes. Additionally, we demonstrate that the semantically enriched information model can facilitate secondary use of clinical data by visualizing the execution of queries based on standard terminologies. Finally, the compatibility of the information model with the IPS standard enables sharing of clinical data among different information models.

Acknowledgement

This research is supported by the National Scientific Program eHealth in Bulgaria.

References

- [1] European Commission, "Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare," Official Journal of the European Union, vol. L88, no. 4.4.2011, pp. 45-65, 1 Jan 2014.
- [2] European Commission, "European Reference Networks," 2017.
- [3] European Reference Network , "Clinical Patient Management System (CPMS)," 2019. [Online]. Available: <https://ern-euro-nmd.eu/clinical-patient-management-system/>.
- [4] European Commission, "European Rare Disease Registry Infrastructure (ERDRI)," 2019. [Online]. Available: <https://eu-rd-platform.jrc.ec.europa.eu/erdri-description>.
- [5] S. Vandeve, A. Elenkova, E. Natchev, G. Kirilov, D. Tcharaktchiev, M. Yaneva and e. a. , "Treatment outcome results from the Bulgarian Acromegaly Database: adjuvant dopamine agonist therapy is efficient in less than one fifth of non-irradiated patients," Experimental and Clinical Endocrinology & Diabetes, vol. 123, no. 01, pp. 66-71, 2015.
- [6] P. Petrossians, A. F. Daly, E. Natchev, L. Maione, K. Blijdorp and M. Sahnoun-Fathallah, "Acromegaly at diagnosis in 3173 patients from the Liège Acromegaly Survey (LAS) Database," Endocrine-related cancer, vol. 24, no. 10, pp. 505-518, 2017.
- [7] CEN/TC 251, "FprEN 17269. Health informatics - The Patient Summary for Unscheduled, Cross- border Care," European Committee for Standardization, Brussels, 2019.
- [8] CEN/TC 251, "FprCEN/TS 17288. Health informatics - The International Patient Summary: Guidance for European Implementation," European Committee for Standardization, Brussels, 2018.
- [9] ISO 13606, "Implementation resources. ISO 13606:2008 XML Schema," 2008.
- [10] E. Siegel and A. Retter, eXist. A NoSQL Document Database and Application Platform, O'Reilly Media, 2014.
- [11] SNOMED-CT, "SNOMED International," 2019. [Online]. Available: <https://www.snomed.org/>.

Data Transfer Model for HIS and Developers Opinions in Turkey

Filiz ISLEYEN ^{a,1}, M. Mahir ULGU ^a

^a *T.C. Ministry of Health, General Directorate of Health Information Systems,
Ankara, Turkey*

Abstract. Hospital Information Systems (HIS) are used in every hospital in our country and these systems have been developed by different vendors. If there is a change in the HIS software used in a hospital for various reasons, all data must be transferred from the current HIS to the new HIS. Accurate, complete, consistent, reliable and rapid transfer of this transfer is important in order not to interrupt the health care provided by the hospital. The aim of this study is to introduce a data transfer model standart (VEM) between independent HISs and to explain the validations that are applied to enforce the model work correctly. In addition, we share the opinions from HIS vendors who develop VEM compatible systems. In the development process of the model, studies with data transfer experts from the HIS vendors in Turkey were performed. A validation software has been developed to check the compatibility of VEM with HIS software. In addition, a questionnaire was sent to HIS developers via e-mail and their opinions on VEM were received. Fourty six HIS developers voluntarily participated in the study, and 70% said they thought VEM helped with solving data transmission problems. As a result, it can be said that VEM has been successful in solving many problems encountered for data exchange in HIS changes in hospitals.

Keywords: Data transfer model, data exchange, HIS.

1. Introduction

In our country and in the world, hospitals, which are the users of Hospital Information System (HIS), could change the system they are using, for the reasons of the negative elements of the important components (modules), usability problems, dissatisfaction, etc. In such cases, hospital managers should prepare for a difficult transition and set realistic expectations [1]. Implementing a new Electronic Health Record (EHR) or switching to another is likely one of the most disruptive predictable events a hospital can experience, affecting practically every employee and workflow at a hospital [2]. Conceptually, most HIS operate on the same type of information (demographics medical history, medications, diagnosis, laboratory findings, billing, etc.). Since medical content standards of HIS were not determined when they were first used in hospitals, most of the HIS' in Turkey have formed their own specific terminology for

¹ Corresponding Author, email: filiz.isleyen@gmail.com

medical data. This unstandardized terminology also was making it difficult to transfer data accurately and quickly in the HIS data transfer process. It may be a tremendous challenge to understand and map the database architecture used in the current HIS to transfer the health history of each patient to a new HIS. Mapping of data at a high level of accuracy, a fully automated transfer while taking into account all the patient information in the hospital is not a big problem in only our country, but also in the world [3]. Accurate mapping of the data into the new system is vital because EHRs serve as the legal health record, containing individually identifiable information describing the healthcare services delivered during a patient's clinic visit [4]. The process of transition from current HIS to new HIS can lead to long, exhausting efforts and data loss [5].

In our country, as the databases to be transferred are different in every HIS and the data transfer process is completed in a longer period than it is anticipated. If there were a universal standard (or at least some kind of "universal translator" for EHR data), data conversion would be much simpler, and the process of switching EHRs would be easier [6]. Given the differences in HIS structures and the diversity of EHRs across countries, it may be premature to speak of a universal model for data transmission. However, a national model independent of the HIS structure may be successful for the data transfer problem. For this reason, a data model has been developed for data transfer and studies have been carried out to use this model as a standard in data transfer between HISs. Minimum Data Transfer Model (VEM) has been developed to be the model that defines the structure and minimum content of any HIS. The minimum here; expresses the data that need to be transferred to the new HIS as a priority.

The aim of this study is to introduce a data transfer model standart (VEM) between independent HISs and to explain the validations that are applied to enforce the model work correctly. In addition, we share the opinions from HIS vendors who develop VEM compatible systems.

2. Method

The model, which was developed as a standard for the use of HIS data transfers, was created by analyzing the database designs of HIS softwares used in the majority of the state hospitals in Turkey. The main purpose was to determine a minimum data transfer model which would cover necessary data elements and which should be able to be transferred primarily between HISs. In this context, a draft model is formed by working together with HIS vendors. VEM is designed as arelational structureusing relational database views. These standard views are created by HIS vendors and reflect the records from the original HIS database. HIS vendors have to follow VEM standards and provide VEM compatible views (HIS-VEM) according to their database design. Because HIS vendors can adopt various software development pattern they are free to desing their own relational database models however they have to provide VEM views on each HIS database system. In some occasions vendors use multiple primary data tables to create certain views and in some occasions VEM to physical database mapping would be one to one. VEM views content is available as

XML at <https://vem.saglik.gov.tr/>. According to the validations made so far, two update studies have been done for VEM. In the first version, the number of views was 63, while in the current version it is 84. All versions of VEM are available at the reference web site.

A validation software (VEM Analysis Program - VAP) has been developed to test the compatibility of VEM with HIS software. VAP was developed using the Delphi programming language in RAD Studio platform and it supports all of the databases (Oracle, MySQL, MsSQL, PostgreSQL etc.) used by HIS software in our country. HIS vendors are required to login to the VAP where they run the database connection information which they use to test the HIS-VEM data transfer model. VAP checks if the data structure and content which are found as XML in <https://vem.saglik.gov.tr/> are compatible with HIS-VEM.

It was seen that the validation process of the each HIS vendors via VAP was taking a long time. For this reason, a workshop was conducted in order to enable HIS developers to work with each other. In the workshop, HIS vendors were paired to work with another HIS vendor which was using same database system. For example, a vendor using Oracle has been paired with a vendor using Oracle. A test form was given to the software developers and they were able to work together for five days by analyzing each other's HIS-VEM views by analyzing the problems that might occur in case of data transfer.

During the HIS-VEM validation, the HIS developers who form the model expressed their opinions on VEM views in written form or verbally. However, in order to get the opinions of the developers who could not give face-to-face interviews because they were busy or they could not join the workshop, a web-based questionnaire was sent to 56 HIS developers via e-mail.

3. Results

The validation process via VAP is done in two steps. In the first step; the views of VEM, which is the standard model and HIS-VEM are compared (data elements in the view, reference codes of the view, SQL script information, comparing the number of records in the view etc.). In the second step; controls related to the quality of the data in the views (PK-Primary Key Control, Reference Table Name Control, Relational Data Control, Time Accuracy Control, Data Accuracy Control, Data Type Control) are performed.

Majority of the HIS vendors tested their VEM compability using the VAP system, however, they all failed the tests initially. Later, we had organized a workshop in which all HIS vendors were invited to attend and at the end of the workshop most the HIS vendors participating to the workshop made successful implementation of standard VEM in to their system and they manage to get compliance with our standards.

A web-based questionnaire was prepared using a 5-point Likert-type scale and sent to 56 HIS developers. Forty-six of them responded. 41 of the participants were male, 5 were female and 94% were university graduates or higher. Their work experience was 12 ± 5 years (mean \pm standard deviation) and the highest working

time was 23 years. 96% of the participants knew about the VEM standard and 91% of them said that they know how to transfer data in the hospitals where they work. The responses of the participants regarding VEM views are shown in Table 1.

Table 1. Opinions of HIS Developers about VEM

Considering the entire systematics generated for VEM, please express your opinion about the following statements.					
Questions	Certainly Disagree (1) n (%)	Disagree (2) n (%)	Undecided (3) n (%)	Agree (4) n (%)	Certainly Agree (5) n (%)
I think that the number of views created in VEM is adequate for the purpose of the developed model.	4 (%8.7)	5 (%10.9)	8 (%17.4)	22 (%47.8)	7 (%15.2)
Views are sufficient to meet field requirements.	3 (%6.5)	7 (%15.2)	12 (%26.1)	14 (%30.4)	10 (%21.7)
I think the explanations of the created views are clear and understandable.	4 (%8.7)	5 (%10.9)	8 (%17.4)	18 (%39.1)	11 (%23.9)
VEM helps to solve problems with data transmission.	3 (%6.5)	4 (%8.7)	7 (%15.2)	15 (%32.6)	17 (%37)
With VEM, it is easier to collect data from the field.	1 (%2.2)	6 (%13)	5 (%10.9)	19 (%41.3)	15 (%32.6)

4. Discussion

Changing HIS can create a host of issues. These include practical issues, such as the sheer amount of time required to research, implement, and transition to a new HIS package; loss of productivity in the process; and changes in functionality from one HIS to another, which require retraining and perhaps may lead to a loss of some functions on which the physicians counted previously [6]. Patient demographics, billing, and other financial data that was stored in the previous HIS will need to be transferred to the new HIS [6]. When the HIS changes in Europe and the United States are examined, it can be seen that the HIS exchange takes a long time and a large part of this period is related to the data transfer part. When the proposed strategies to conclude the changes in electronic health record in the easiest way are examined, it is seen that the importance of using standard terminology of different HISs is emphasized [3].

The developed model has been used by various HIS companies in data transfer since 2015 in Turkey. The aim of the model is to determine the data (patient information, laboratory results, diagnostic information, etc.) that should be transferred in order to prevent disruption of the health care processes in the hospitals in order to ensure that these data are transferred accurately, completely and consistently. Therefore, the model is named as the minimum data model. The feedback received from the HIS developers using the model has shown that VEM has been successful in

data that needs to be transferred primarily. However, when all the data in the hospital need to be transferred from the existing HIS to the new HIS, it was seen that the HIS developers use traditional methods in addition to VEM. This necessitated the transfer process to be performed in two different ways and caused the HIS developers to perform two separate tasks. According to our experience of VEM validation study or feedbacks from HIS developers using the model, it was seen that the developed model-VEM gave faster results compared to the traditional method, but a model that would provide all the data to be transferred in a hospital would be more beneficial. For this reason, new studies have been planned to improve the scope of the VEM.

Due to the success of many HIS vendors in our workshop, workshops create an effective learning environment for HIS developers. We also believe that the collaboration of HIS developers and their successful results increased the adoption of VEM.

In Table 1, 70% of respondents responded positively to the question “VEM helps to solve data transfer problems, 15% stated that they did not agree and 15% stated that they were undecided. The majority of the participants think that the model is successful compared to the traditional method.

Although VEM is a standard that was developed for data transfer, it has been learned that most HIS software developers have made changes to their database structure after VEM. It is considered that VEM shows a positive effect on data standardization.

5. Conclusion

As a result, the majority of HIS developers stated that VEM solves the problems in data transfer. In light of this information, it can be said that the model developed to be used as a standard in HIS changes in hospitals has been successful for the targeted aim. This fact also revealed the necessity of extending the model to a wider spectrum of the data in any hospital. By expanding the VEM, data transfer processes in hospitals in our country can be completed more quickly and smoothly.

References

- [1] J.J. Saleem, J. Herout, Transitioning from one Electronic Health Record (EHR) to Another: A Narrative Literature Review. *Proceedings of the Human Factors and Ergonomics Society 2018 Annual Meeting*, (2018), 489-493.
- [2] M. Barnett, A. Mehrotra, A.B. Jena, Adverse inpatient outcomes during the transition to a new electronic health record system: observational study. *BMJ*, (2016) 354 :i3835.
- [3] Electronic Health Reporter, 10 Tips to Successful Data Migration During an EHR Change, <http://electronichealthreporter.com/eight-tips-successful-data-migration-ehr-change/>. Accessed: 26.12.2019.
- [4] D. Dolezel, J. Moczygemba, Implementing EHRs: An Exploratory Study to Examine Current Practices in Migrating Physician Practice. *Perspectives in Health Information Management*. (2015), 12.
- [5] R.L. Edsall, K.G. Adler, EHR Switch Survey: Responses From 305 Family Physicians. *Family Practise Management* 22(1), (2015) 13-18.
- [6] D. F. Shay, Physicians Switching EHRs, *Health Law Handbook*, Thomson Reuters, Philadelphia, 2018.

Decision Quality Is a Preference-Sensitive Formative Concept: How Do Some Existing Measures Compare?

Jack DOWIE^{ab1}, Mette Kjer KALTOFT^b,

^{ab}*London School of Hygiene and Tropical Medicine, UK*

^b*University of Southern Denmark, Denmark*

Abstract. The primary output of a decision making process is a decision and a key outcome measure is therefore decision quality. However, being a formative construct, ‘decision quality’ is both preference- and context-sensitive and legitimate alternative measures accordingly exist. A decision maker wishing to measure decision quality in the evaluation of a decision or decision making process needs to be aware of the attributes of the measures on offer. This paper establishes some of the key conceptual differences by examining two measures: Decision Quality Instruments and MyDecisionQuality. Four of their main conceptual differences relate to: the timing of the measurement (at the point of decision or at follow-up when the ‘downstream’ outcome is known); (whether or not an objective assessment of the information state of the individual is included (as opposed to self-reported state); whether the instrument itself is preference-sensitive; and whether the measure is to be used in the context of individualised clinical practice at the point of care or only in research to produce group level feedback. Establishing agreed measures of decision quality is necessary and useful, so long as it is accepted that it is a preference- and context-sensitive construct, in the way that is widely acknowledged in relation to, for example, Health-Related Quality of Life, with its many measures.

Keywords: decision quality, Decision Quality Instruments, MyDecisionQuality, formative construct, preference-sensitive.

1. Introduction

If the fundamental aim of informatics is to help decision makers make high-quality decisions, it might be expected that the measurement of ‘decision quality’ would be a major focus of interest. However, there is limited attention to this *outcome* in the literature, compared to the amount devoted to related *input* qualities: the quality of the information inputs being produced, the quality of the transmission and dissemination of information inputs, and the quality of the processing of the information inputs within decision making. In this paper we explore two specific approaches to the measurement of decision quality, an exploration which, while restricted in scope, is sufficient to throw significant light on the reasons for this relative paucity of attention. Among other things, it may help explain the surprising non-appearance of decision quality as a

¹ Corresponding author: Jack Dowie, LSHTM, 15-17 Tavistock Place, London, UK WC1H 9SH; email: jack.dowie@lshtm.ac.uk

Patient-Outcome Measure (PROM). In the comprehensive 2015 survey of the latter [1], the term ‘decision quality’ appears only within the section headed ‘Patient experience’ and not as a PROM.

Among the underlying reasons for this marginalisation is that the latent unobservable construct (‘decision quality’) is a *formative*, not a *reflective* construct, and therefore preference- and context-sensitive [2]. This has both methodological and practical consequences. Among the methodological consequences are the psychometric standards appropriate for validating instruments seeking to measure it. ‘Decision quality’ is an ‘it’ that does not exist in the sense that a reflective construct (such as appendicitis) exists, where psychometric standards such as internal consistency are valid. The main practical consequence of its formative nature is that there are many possible and acceptable definitions of decision quality, none of which can be validated as the ‘gold standard’. The ambition to develop a universally agreed definition of a formative construct is understandable since widely-agreed definitions have their uses and in many senses are essential. The various formative measures of Health-Related Quality of Life exemplify acceptance of their intersubjective, value-based provenance. However, capitalising a construct as Shared Decision Making or Decision Quality cannot produce ontological transformation from formative to reflective, even if it is sometimes institutionally convenient to treat it as if it does.

There are many meaningful distinctions which can and should be made between different proposed measures of decision quality, and the basic point being made here is that the *decision* on which measure of decision quality to select for use, from a set of alternative measures given this label, is both *preference-sensitive* and *context-sensitive*. *Different instruments will measure different constructs and are not to be regarded as different measures of the same construct.*

A decision maker selecting the measure that is optimal for the particular evaluation they are undertaking needs to be informed about their content and construct validity. The limited aim here is to establish *some* of the key *conceptual* differences in measures of decision quality. We set aside most practical differences, the need for alternative versions or adaptations for contexts such as low literacy or ethnicity being taken for granted. Examining two measures will be sufficient for our limited purpose: the Decision Quality Instruments (DQIs) [3-4] and MyDecisionQuality (MDQ) [5-6]. The concept is often used implicitly as an umbrella for a set of individual indicators [7].

2. The Decision Quality Instruments (DQIs)

A generic DQI framework is used in the production of condition-specific instruments, 14 of which are currently available, including ones for breast cancer surgery, prostate cancer screening and treatment, back pain and knee and hip osteoarthritis. (<https://mghdecisionsciences.org/tools-training/decision-quality-instruments/>)

Since the DQIs are condition-specific, we will focus on that for Breast Cancer Surgery (BCS-DQI) [3-4]. The BCS-DQI comprises a set of 15 ‘objective’ knowledge questions ‘to determine whether patients are informed’ and a set of 8 questions about their ‘goals and concerns’. As a profile measure, it produces two scores. The *DQI-Knowledge Score* is the number of correct responses divided by the number of items, resulting in scores from 0% to 100%. A threshold for considering a patient to be ‘well-informed’ is set, using (if available) the mean knowledge score for a group of patients who have viewed a decision aid. The *DQI-Concordance Score* measures ‘the extent to

which patients received treatments that reflected what is most important to them'. It is derived from a multivariate logistic regression model in which the treatment received (e.g. surgery vs. non-surgical) is the dependent variable. The goals and concerns specified by the respondent are the independent variables, with other factors that should influence treatments (e.g. stage of disease for breast cancer surgery) included, as needed. The dependent variable is a predicted probability of treatment for each patient. Patients with a predicted probability of more than 50% who had the treatment, and those with a predicted probability equal to or less than 50% who did not have the treatment, are classified as having treatments that match their goals and are 'concordant'. (Dividing the number who matched by the total number in the sample then yields a summary percentage concordance score for the group.) A binary *Decision Quality Composite Score* can be created with a score of 1 for patients who were well-informed [i.e. met the pass threshold] and received treatments matching their preferences, and 0 for all others.

The interpretation of both the 2-score profile and the binary composite score is only at the group level and available only after follow-up, months after the decision. So the DQIs are essentially research tools for evaluating a decision process (which is assumed to be 'shared decision making', usually including a decision aid), for the purpose of improving future practice through group level feedback. They are not envisaged as tools to be used as the source of individual feedback in real time within clinical practice.

3. MyDecisionQuality (MDQ)

MDQ is a preference-sensitive instrument in which individuals provide their responses as soon as possible after a decision is made. The responses are of two distinct types - their personal ratings of the preceding decision- and decision-making process on 8 criteria, and their weightings for those 8 criteria. The MDQ score is the weighted sum of the ratings (i.e. the expected value for each option). The abbreviated labels for the 8 items are four which relate to the decision content - Options, Effects, Criterion Importance, and Effect Chances - and four which relate to the decision process - Trust, Support, Control, and Commitment [5-6].

MDQ is a self-reported, preference-sensitive ('dually personalised') multi-criterial index, generic across both conditions and contexts. Its primary purpose is to evaluate the quality of a decision and decision process and do so at the individual level. The evaluation is to occur as soon as possible after the decision is taken. This is to rule out reference to any downstream outcomes, for whatever reason. (Two common ones are to establish whether action or behaviour subsequent to the decision was in accordance with the person's expressed 'goals and values' and to permit their *experienced* regret to affect the assessment of decision quality.) It also excludes 'objective knowledge possession' as a criterion. However, it addresses both the decision and the decision process, being designed to indicate whether the 'reasonable patient' legal standard for *informed and preference-based* consent has been met (items 1-4), as well as the 'subjective patient' ethical standard for the care process (items 5-8). Where clinicians agree to complete an equivalent measure, it can provide decomposed measures of concordance (e.g. between patient and practitioner) and so establish which ratings or weightings are the source of any discordance. A sub-analysis indicates the criteria which hold the greatest potential for improved future decision quality. It is possible to

aggregate MDQ scores across individuals and arms of trials [8], but the interpretation of these results must be approached with caution, given their individual preference-sensitivity.

4. Discussion

A wide variety of constructs of decision quality are possible, with the main distinctions resting on *why* they are done, *who* is asked to provide *which* type of responses/s to *what* items, and *when* and *where* they are asked to do so. The answers determine whether the measure of decision quality is preference-sensitive in relation to the items in the instrument; whether the measure is useful in real time clinical practice or is essentially an output from evaluation research at a group level; and whether it is appropriate in evaluations of a decision, of a decision process, or a decision aid (including policy makers and regulators approving the last of these).

The timing of measurement is perhaps most important, because for many it impinges on the validity of the measure, as well as its use. Elwyn and Miron-Shatz report the widespread view that 'Good (or bad) decisions may have good or bad outcomes by virtue of chance, at the level of individual consequences.... post hoc assessments, based on the outcomes of decisions, are unsafe measures of good decisions.' [9] (p141-2). Delayed assessment also permits regret to enter into any evaluation at 'downstream' follow-up. *Experienced* regret - indeed all hindsight - is a valuable input into a future decision, but only if it is not allowed to affect the evaluation of a previous one, in which one has had the opportunity to input *anticipated* regret, in foresight. Dissonance reduction is another post hoc influence on evaluations of decision quality widely regarded as undesirable.

Incorporation of an objective knowledge test is a further example of how a different construct of decision quality is created. Decision makers in some contexts may prefer a measure reflecting responses from 'objectively informed persons'. Others will reject this: '... knowledge as a necessary component of a good decision - both as an evaluation of the effectiveness of deliberation and as an evaluation of the determined decision - may be difficult to operationalize and, at a deeper level, may reflect an untested assumption - that a standardized, pre-specified level of knowledge is necessary in order to achieve a high level of deliberation.' [9] (p144).

5. Conclusion

Since the latent construct of decision quality is formative, each of these instruments measures the construct it measures. An instrument should be chosen for use in an evaluation if its characteristics are preferred and it is context-relevant, as deemed by the decision maker. Ambitions to produce a standardised, preference- and context-free decision quality assessment instrument should be resisted, with the different products accepted as a menu from which a preference- and context-sensitive selection can be made. We note that it is claimed that the definition of decision quality in DQIs has been validated by the International Patient Decision Aid Standards group [4,10]. While this establishes its serious credentials as a construction of decision quality, as with any formative construct these rest on the intersubjective agreement about its contents (items and weightings), reflecting the beliefs and preferences of the producing group. MDQ

does not seek to answer the philosophical question of what constitutes a ‘good decision’ [11], merely to provide a measure of ‘decision quality’ with its particular set of characteristics. Decision quality should be regarded as *the* fundamental outcome of a decision making process, despite the challenges its formative character presents.

Funding

The software used in current implementations of MDQ was installed at <https://ale.rsyd.dk> in a nationally funded project to develop decision support tools. SATS J.nr.1-1010/116/27.

Conflict of Interest

Jack Dowie has a financial interest in the commercial use of the Annalisa software.

References

- [1] D. Cella, E. Hahn, S. Jensen, Z. Butt, C. Nowinski, N. Rothrock, et al., *Patient-Reported Outcomes In Performance Measurement. Research*, Triangle Park (NC); RTI Press 2015.
- [2] M.K. Kaltoft, J.B. Nielsen, J. Dowie, Formative preference-sensitive measures are needed in person-centred healthcare at both clinical and policy levels, *European Journal for Person Centered Healthcare* **5** (2017), 17–20.
- [3] K. Sepucha, E. Ozanne, K. Silvia, A. Partridge, A. Mulley, An approach to measuring the quality of breast cancer decisions, *Patient Education and Counseling* **65** (2007), 261–9.
- [4] K. Sepucha, J. Belkora, Y. Chang, C. Cosenza, C. Levin, B. Moy, et al., Measuring decision quality: psychometric evaluation of a new instrument for breast cancer surgery, *BMC Medical Informatics and Decision Making* **12** (2012), 51.
- [5] M.K. Kaltoft, M. Cunich, G. Salkeld, J. Dowie, Assessing decision quality in patient-centred care requires a preference-sensitive measure, *Journal of Health Services Research and Policy* **19** (2014), 110–7.
- [6] M.K. Kaltoft, J.B. Nielsen, G. Salkeld, J. Dowie, Enhancing healthcare provider feedback and personal health literacy: dual use of a Decision Quality measure, *Studies in Health Technology and Informatics* **218** (2015), 74–9.
- [7] K. McAlpine, K. Lewis, L. Trevena, D. Stacey, What is the effectiveness of Patient Decision Aids for cancer-related decisions? A systematic review subanalysis, *JCO Clinical Cancer Informatics* **2** (2018), 1–13.
- [8] G. Salkeld, M. Cunich, J. Dowie, K. Howard, M. Patel, G. Mann, et al., The role of personalised choice in decision support: A randomized controlled trial of an online decision aid for prostate cancer screening, *PLoS One* **11** (2016), e0152999.
- [9] G. Elwyn, T. Miron-Shatz, Deliberation before determination: the definition and evaluation of good decision making, *Health Expectations* **13** (2010), 139–47.
- [10] G. Elwyn, A. O'Connor, D. Stacey, R. Volk, A. Edwards, A. Coulter, et al., Developing a quality criteria framework for patient decision aids: online international Delphi consensus process, *BMJ* **333** (2006) 417–9.
- [11] J. Hamilton, S. Lillie, D. Alden, L. Scherer, M. Oser, C. Rini, et al., What is a good medical decision? A research agenda guided by perspectives from multiple stakeholders, *Journal of Behavioral Medicine* **40** (2017), 52–68.

Defining the Contextual Problem List

John MEREDITH^{a,1}, Ian MCNICOLL^b, Nik WHITEHEAD^c, Kemi ADEMOYE^c

^a *Wales Institute for Digital Information, Cardiff UK*

^b *OpenEHR International, London UK*

^c *University of Wales Trinity Saint David, Swansea UK*

Abstract. The Problem Oriented Medical Record (POMR) is considered a key charting method to support clinical care. Although not uniformly represented amongst digital health systems, this paper presents a clinical model to represent multiple clinical perspectives from a single problem list. The contextual problem list model is defined according to primary diagnosis, comorbidities and problems arising from the primary condition. It is represented within the patient record as a single composition according to the prescribed context. The model pattern could help alleviate the traditional criticisms of paper and digitally based problem records.

Keywords. Contextual problem list, problem oriented medical record, problem list, clinical models, digital health, electronic health records.

1. Introduction

Health records are traditionally ordered into individual data silos such as pathology, medications or radiology. One such silo, the problem list is represented within the digital domain with variable degree of functionality[1] despite the rules of the concept being well understood[2]. Medical care pathways are traditionally arranged to support the management of multiple problems in the primary care setting through to secondary care via a range of diagnostic procedures[3]. This iterative process achieves a more granular focus with increasingly detailed problems for specific conditions such as diabetes.

The Welsh Information Solution for Diabetes Management (WISDM) is being developed as part of a national digital platform. It is *by definition* a problem oriented medical record (POMR). Its specification describes functions to support the curation of a problem list according whether an entry is a complication of the primary disease, or other important comorbidity that may affect ongoing treatment. This paper describes a clinical model for a contextual problem list (CPL) as a persistent composition within the electronic health record (EHR) to support this requirement.

2. Background

As suggested by Dr Lawrence Weed, there is a need to align the patient record according to the problem[4]. Since the first computerised systems began to appear there has been great variability in how the problem list manifests in the (EHR). Merely replicating paper-based processes and transposing them to the digital domain are not sufficient.

¹ John Meredith, E-mail: john.meredith@wales.nhs.uk.

Manual curation of a list is inherently burdensome and data quality impacts how likely it is to be used and updated[5]. The rules on what should or should not be included in the problem list are less clear[6]. As a result, the problem list as a component of the wider digital record has been shown to not always be available or necessarily complete[7,8].

Disadvantages of the POMR include emphasis of when problems manifest opposed to their priority, and the repetitious nature of their structure[9]. The effect of “compartmentalisation” in inhibiting the “synthesis” of multiple problems and the additional risk of substantiating a differential or incorrect diagnosis as the clinical truth is also cited[10]. POMR can be seen as a barrier to the patient narrative through the propagation of increasingly granular but clinically similar problems[1]. This is further complicated by the need for different views depending on clinical perspective or role and the nature of problems being recurrent[11]. Despite this, the problem list is a persistent function within many modern EHR systems.

Weed[4] originally envisaged his chart as a flat list of entries marked as active or inactive with nested variants where appropriate. This juxtaposed the common pattern found within secondary care reporting or outpatient letters of primary problem/ diagnosis with comorbidities. The latter focused on a single problem or care pathway while the Weed POMR takes a more holistic approach.

Grouping individual entries as a diagnosis or comorbidity is key for billing and the attribution of expended resource. However, it might be postulated that it provides direct clinical value in prioritising and focussing a set of complex clinical issues around the presented problem (and perhaps more so than a POMR styled list). We therefore present the pattern defined within this paper for the CPL as a hybrid of both approaches.

3. Definitions

Within the context of this research, a “problem” may be a presenting complaint or confirmed diagnosis. The nature of the CPL means that the following definitions are inherently flexible and individual records may exist within any given category depending on the clinical perspective and may change over time. As a result, the underlying model needs to support the ability to refute, dispute and replace individual entries and for those changes to be clear and visible within the patient record. Utilising coded entries derived from a terminology service is preferred although this represents a point of note for implementers to observe and is not mandatory.

Primary diagnosis represents the condition for which the patient is receiving care and frames the clinical context of the CPL. While the primary diagnosis may be a variation of a range of entries, these should be limited to a specific care pathway (in this case diabetes). It is possible to initiate a CPL without a primary diagnosis being present as the context itself is categorised within the associated metadata of the composition.

Co-morbidity is a coded diagnosis that does not require management or direct acknowledgment by clinicians within a specified disease domain. However, it is of clinical importance to the ongoing care of the specific disease context. It can be assumed that comorbidity entries will be actively managed through alternate care pathways, but they are compatible with the underlying information model for primary diagnosis.

Problems and complication records represent a diverse range of entries. They may contain a clinical problem that has arisen from the primary diagnosis, is being investigated as a result of the problem or is a keynote for the clinician to consider such as a recent bereavement or learning disability. Disease complications will require active

management by the clinician and maybe the result of referrals from other care settings. Entries within this section may reside from clinical findings (e.g. a presenting complaint of pedal oedema) or disease related disorders (e.g. diabetic retinopathy).

The final category concerns entries that may derive from a variety of findings or disorders, but are not of direct concern or relate to the context instantiated with the CPL. However, these records cannot be ignored as they may become important in future. Consideration must be given by the clinician to reconcile entries from a variety of sources.

The approach defined here means that a diabetologist may be treating a patient with some form of diabetes and record a primary diagnosis of Type I diabetes mellitus with peripheral angiopathy. Conversely, a renal physician could be treating a patient with a primary diagnosis of chronic kidney disease. Both primary diagnoses are dependent on the clinical context, and in some cases may be linked if one disorder is predicated by the other (e.g. chronic kidney disease due to type 2 diabetes mellitus). Both entries are also suitable for inclusion under the comorbidity heading of their respective CPL.

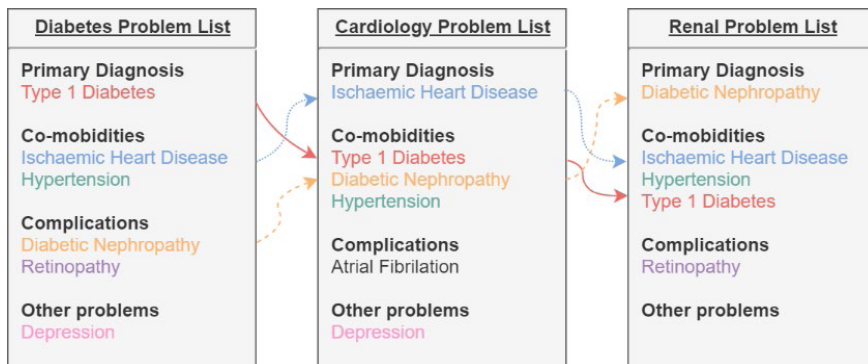


Figure 1. Shared problem records within three contextual problem lists.

A diagnosis of “diabetic nephropathy” may be considered a complication of the patient’s primary diagnosis of Type I Diabetes i.e. this condition would not have occurred had the patient not had this primary diagnosis. However, this could also be considered the primary diagnosis for a nephrologist (Figure 1).

3.1. Model Development

The CPL was modelled with openEHR Archetype Designer[12]. Figure 2 describes the applied model pattern. Each context category is embedded within the relevant section with additional attributes from problem/diagnostic qualifier to support, barring the “Other problem”. This is to ensure that the pattern cannot be disrupted through incorrect attribution of problem categories within the “Other Problem” section.

Additional data elements have been included to support the wider provenance of the CPL including an absence statement that describes why the list may be empty. For example, a patient early on the diagnostic pathway may be considered to have a finding of pre-diabetes. It may not be appropriate to record this as a primary diagnosis although it is still a condition of note. A recorded entry that there are “no known problems” or “no primary diagnosis” is a potentially valuable statement of fact for a patient at this stage.

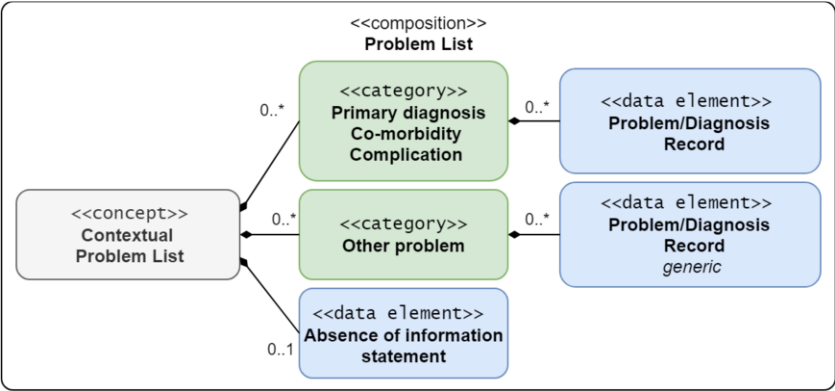


Figure 2. Information model for the contextual problem list template. Each category is optional.

4. Discussion

While the criticisms of POMR are legitimate when concerned with paper, digital functionality can support or eliminate many of them. Duplication of data, a concern in the paper domain[2], is mediated where a digital repository is able to act as a single source of truth. Orientation of the record around the problem, as Weed suggested, is key to facilitating benefits such as more thorough and increasingly relevant medical charts with faster information retrieval[1]. By utilising the CPL as a digital composition, it is possible to view a problem list with a specific clinical perspective such as diabetes as well as a master list of all other problem records. This master list is considered a core component of the EHR[13].

The design of the CPL outlined here seeks to simplify the wider requirement for a single instance of the POMR by creating individual compositions that relate to specific problems or conditions. This is analogous to the “index concept” approach[14]. The level of granularity is focussed upon a primary condition and the many subproblems that may be present with complex cases.

Where accuracy of the problem list is high, clinical errors can be reduced and patient safety increased [15]. This pattern facilitates multiple actors in the clinical pathway to create, review and update problems using data that is recorded once but shared widely. Data quality is of specific concern if we are to rely on the digital record. There is evidence that completeness of data is encouraged when alerting functions are presented to the user[16]. It has been suggested that there are benefits for sharing the POMR in cases of chronic disease management spanning multiple care domains[17]. In addition, this defines a basis from where algorithm-based care can be provided via well defined, standards-based data models.

5. Conclusion

The proposed method for representing the CPL can support multiple care domains from within a single data repository. It represents a hybrid of the traditional POMR and context specific records. While not a definitive list of all patient problems, it presents the key issues that are of concern for the management of the patient’s condition at a point in time.

Individual entries within a CPL may be recorded by other clinicians and considered of interest by other clinical specialties for inclusion in their own contextual problem lists.

References

- [1] S.M. Chowdhry, R.G. Mishuris, and D. Mann, Problem-oriented charting: A review, *Int. J. Med. Inform.* **103** (2017) 95–102. doi:10.1016/j.ijmedinf.2017.04.016.
- [2] A.R. Feinstein, The problems of the “problem oriented medical record,” *Ann. Intern. Med.* **78** (1973) 751–762. doi:10.7326/0003-4819-78-5-751.
- [3] E.P. Balogh, B.T. Miller, J.R. Ball, C. on D.E. in H. Care, B. on H.C. Services, I. of Medicine, and E. and M. The National Academies of Sciences, The Diagnostic Process, (2015).
- [4] L.L. Weed, Medical Records That Guide and Teach, *N. Engl. J. Med.* **278** (1968) 593–600. doi:10.1056/NEJM196803142781105.
- [5] C. Holmes, M. Brown, D. St Hilaire, and A. Wright, Healthcare provider attitudes towards the problem list in an electronic health record: a mixed-methods qualitative study, 2012. doi:10.1186/1472-6947-12-127.
- [6] A. Wright, F.L. Maloney, and J.C. Feblowitz, Clinician attitudes toward and use of electronic problem lists: a thematic analysis, *BMC Med. Inform. Decis. Mak.* **11** (2011) 36. doi:10.1186/1472-6947-11-36.
- [7] R. Mann, and J. Williams, Standards in medical record keeping, *Clin. Med. (Northfield. Il).* **3** (2003) 329–332. doi:10.7861/clinmedicine.3-4-329.
- [8] A. Wright, A.B. McCoy, T.T.T. Hickman, D.S. Hilaire, D. Borbolla, W.A. Bowes, W.G. Dixon, D.A. Dorr, M. Krall, S. Malholtra, D.W. Bates, and D.F. Sittig, Problem list completeness in electronic health records: A multi-site study and assessment of success factors, *Int. J. Med. Inform.* **84** (2015) 784–790. doi:10.1016/j.ijmedinf.2015.06.011.
- [9] Lippincott Williams & Wilkins., Complete guide to documentation., Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.
- [10] S.E. Goldfinger, The Problem-Oriented Record: A Critique from a Believer, *N. Engl. J. Med.* **288** (1973) 606–608. doi:10.1056/NEJM197303222881205.
- [11] P. Salmon, A. Rappaport, M. Bainbridge, G. Hayes, and J. Williams, Taking the problem oriented medical record forward., *Proc. AMIA Annu. Fall Symp.* (1996) 463. <http://www.ncbi.nlm.nih.gov/pubmed/8947709>.
- [12] EhrScape Archetype Designer, (2019). <https://ehrscape.marand.si/designerv2/>.
- [13] R.S. Dick, E.B. Steen, and D.E. Detmer, eds., The Computer-Based Patient Record, National Academies Press, Washington, D.C., 1997. doi:10.17226/5306.
- [14] D. Clercq, The Index as a New Concept towards an Integrated Framework for the Electronic Patient Record, *Methods Inf. Med.* **41** (2002) 313–320. doi:10.1055/s-0038-1634388.
- [15] W.L. Galanter, D.B. Hier, C. Jao, and D. Sarne, Computerized physician order entry of medications and clinical decision support can improve problem list documentation compliance, *Int. J. Med. Inform.* **79** (2010) 332–338. doi:10.1016/j.ijmedinf.2008.05.005.
- [16] C. Wright, J. Pang, J.C. Feblowitz, F.L. Maloney, A.R. Wilcox, K. Sax McLoughlin, H. Ramelson, L. Schneider, D.W. Bates, and A. Wright, Improving Completeness of Electronic Problem Lists through Clinical Decision Support, *J. Am. Med. Informatics Assoc. JAMIA.* **19** (2016) 555–561. doi:10.1136/amiainl-2011-000521.
- [17] L. Samal, A. Wright, B.T. Wong, J.A. Linder, and D.W. Bates, Leveraging electronic health records to support chronic disease management: the need for temporal data views., *Inform. Prim. Care.* **19** (2011) 65–74. <http://www.ncbi.nlm.nih.gov/pubmed/22417816> (accessed September 19, 2019).

Digital Follow-Up Application for Cancer Patients – Value Mechanisms Regarding Health Professionals’ Work

Sauli MIETTINEN^a and Henni TENHUNEN^{a,1}

^a*Institute of Healthcare Engineering, Management and Architecture (HEMA Institute),
Department of Industrial Engineering and Management, Aalto University, Espoo,
Finland*

Abstract. This study examined the value mechanisms of a digital cancer follow-up application (CFUA). Value is defined as the health outcomes achieved per resources used. The focus was on how the digital health intervention (DHI) affects the professionals’ work. The research was conducted through interviews of healthcare professionals in the empirical context of gynecological cancer clinic at a university hospital in Finland. The identified key mechanisms were improved coordination and optimized care level. The CFUA improves coordination through reducing time restrictions of work, as the professionals have more chances to time their contacts and plan their work. The CFUA helps in optimizing care level, as the application can effectively filter patients to either nurses or physicians according to symptoms. Coordination mechanism is closely connected to other healthcare value mechanisms, such as reducing waste and supply management. This study increases understanding of DHI value potential recognition, and how value can be actualized.

Keywords. Telemedicine, Cancer Care, Value-Based Healthcare

1. Introduction

Introducing new digital health interventions (DHIs) requires that their cost-effectiveness, their value, is proven. Implementing DHIs with demonstrated value improves the economic sustainability of the healthcare system and benefits all parties, including patients, payers, providers and suppliers [1]. Porter defines value as the patient health outcomes relative to the costs of care for the patient [1].

Cancer care DHIs have been a subject for academic research. It has been found that digital interventions can through certain mechanisms help cancer patients to cope better with the disease and symptoms [2]. Patient-reported-outcome follow-up has been shown to improve cancer patients’ survival and quality of life [3,4]. According to [5], telemedicine in cancer care can reduce the travel costs for patients and thus improve total value. Web-based outcome surveillance was found to be a cost-efficient way of monitoring and can reduce the follow-up costs for the insurance-provider [6]. It has been stated that also the staff viewpoint should be considered in digital services in cancer care [7]. Thus, when analyzing an intervention’s potential for improving value-based healthcare, it is important to examine the personnel perspective, where a

¹ Corresponding Author: E-mail: henni.tenhunen@aalto.fi

significant part of the costs are incurred. Therefore, this research took intervention's impact on healthcare professionals' work as the focus and aimed to find the mechanisms that cause changes in their behavior and affect costs.

The studied DHI is a modular digital cancer follow-up application (CFUA) called Kaiku Health. This study focuses on the use of CFUA for gynecological cancer patients. The context of the CFUA case is the gynecology unit of a tertiary hospital in Finland. In the unit, all types of gynecological cancers are treated. CFUA is used both during treatments and in monitoring after treatments. The people using CFUA include patients, specialist physicians and nurses, some of which work with patients during cancer treatments, and some with patients in monitoring after treatments. All specialist physicians are specialized in gynaecology, and their tasks include diagnosis and treatment of gynecological cancer patients. Most of the patients are over 65 years old. The patients use CFUA to report their health status and symptoms by filling symptoms and quality of life questionnaires and can contact healthcare personnel through CFUA with chat messages. The message feature also allows sharing files, such as medical certificates. When a member of the personnel logs in, CFUA shows the patient cases that have new information that should be handled. Healthcare professionals react to the contacts either through CFUA or in acute situations by phone (Figure 1).

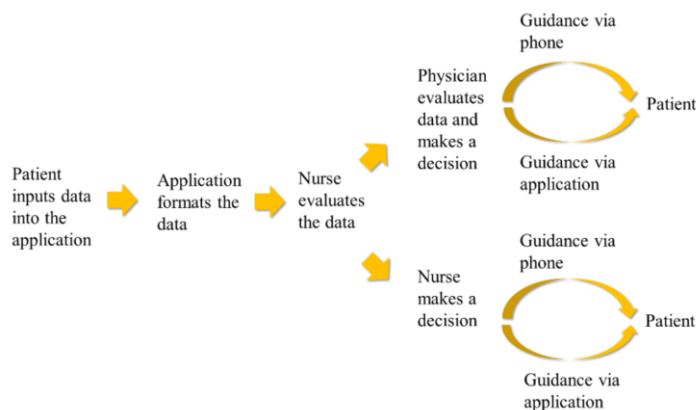


Figure 1. Typical data flows in the use of CFUA.

The application works with mobile devices and desktop devices. CFUA is offered as a separate installation for every customer clinic [7]. In WHO's (2018) classification of digital health interventions, CFUA can be described as a telemedicine intervention including e.g. consultations between remote client and healthcare provider, and remote health monitoring [8]. CFUA also possesses features of personal health tracking interventions, as it allows self-monitoring of health in form of symptom data.

The objective of this study was to identify and analyze the mechanisms through which CFUA creates value in cancer care from the healthcare professionals' perspective in the context of tertiary gynecological cancer clinic. The research problem was approached using the CIMO logic, describing the context, intervention and outcomes to analyze the mechanisms [9,10]. CFUA's CIMO-configuration and the changes were examined. We posed the following research question: How does a digital follow-up application change the work of healthcare professionals and which value mechanisms are activated through these changes?

2. Data and Methods

This research was conducted as a single case study. The unit of analysis was the work of a healthcare professional, which comprises tasks, actions and work processes, as processes can be complemented with other units of analysis, especially when they are not always clearly manifested [11]. The data was collected through semi-structured interviews with eight members of the clinic’s healthcare professionals (N=8). Four informants were specialist physicians and four informants were registered nurses. Three of the nurses worked with patients during the cancer treatments and one nurse worked with patients after the treatments (Table 1). The interview data was transcribed verbatim, and sorted into categories.

Table 1. List of informants.

ID	Role in Organization and Cancer Treatments
Nurse 1	Works with patients in active treatment
Nurse 2	Works with patients in diagnostics and monitoring
Nurse 3	
Nurse 4	Works with patients in active treatment
Physician 1	Specialist, diagnostics and treatment of gynaecological cancers
Physician 2	
Physician 3	
Physician 4	

The data collected from the interviews was analyzed qualitatively, using the Grounded Theory methodology. From the data, core categories were formed that described the relevant themes summarizing the data set [12].

3. Results

The identified value mechanisms were improved coordination, right care delivery level and timing, reduction of waste, demand/supply-management, control information and health co-creation through symptom-based care (Table 2). Especially the mechanism of coordination comprises sub-mechanisms, such as reduced variation and is connected to other mechanisms.

Having more accurate and exact data about the patient enables optimizing the care level. CFUA provides data about the patients’ symptoms, which is used to decide whether a physician or nurse should handle the case. Nurses first read the non-urgent CFUA contacts, and only forward them to the physicians if necessary. This allows a more efficient care level, as doctors only handle more challenging cases fitting their skills.

Variation in tasks hinders productivity, as mental preparation and adaptation are needed when changing between tasks. The health professionals argued that they are able to follow their schedules better with CFUA, as the unpredictable phone calls do not interrupt and interfere with their current work. CFUA allows scheduling the day into sub-parts where the focus stays on a certain type of task for a longer period of time. Reduced time dependency allows the professionals to perform tasks with more efficient scheduling. The coordination mechanism of CFUA is primarily connected to organizing patient care and delivery of services in a timely manner. Instead of facilitating the organization of cancer care in a system level between teams and providers, CFUA is a tool for coordinating cancer care between the patient and specific nurses and physicians assigned to the patient.

Table 2. Value mechanisms of the CFUA.

Mechanism	Summary
Coordination	CFUA allows individual professionals to work more flexibly and plan their work better, reducing interruptions and improving productivity.
Right timing (for personnel)	With CFUA, the professionals have a time-window that allows them to choose a suitable time to handle a contact.
Care delivery level	CFUA allows optimizing care delivery to the lowest level by directing only the most challenging contacts to physicians.
Reduction in waste	As CFUA reduces interruptions in professionals' work, they need to spend less time.
Control information	CFUA provides more complete information, such as time series of symptoms, to professionals. This information helps professionals to decide what they should do next, thus improving coordination of care.
Demand/supply management	By relieving the time restriction of communication, CFUA allows scheduling the supply, the answers to contacts, more flexibly.
Symptom-based care	Patients contribute to their care by actively monitoring and reporting their health status. With reduced routine appointments, the professionals focus more on providing care for patients who report symptoms.

Improved control information can lead to a more optimal processing as the right decision can be made. The professionals using CFUA stated that the time-series of the patients' earlier symptoms help in comparing the situation with previous weeks, and thus evaluating if further actions are needed. Reducing waste "Muda" happens when improved coordination of work reduces non-value adding tasks. In addition to providing more data about patients, the CFUA can also provide more structure to monitoring processes. With CFUA, the processes can potentially be more standardized, if the contacts are more predictable. On the other hand, any lack of clarity in the procedures of CFUA use can cause waste.

In addition to the CFUA, the demand in the clinic is actualized through two channels: phone calls (scheduled and non-scheduled) and clinic visits. Although the timing of demand remains roughly the same, the channel of demand is connected to the timing of supply. In phone calls, the professionals must answer when a patient calls, whereas for CFUA contacts they have up to two days' time to answer. Supply can be planned in a more flexible way, while still retaining sufficient patient experience, as the patients are instructed not to use CFUA for urgent matters. The mechanism of demand/supply-management is related to the coordination mechanism. The ability to time supply with a larger degree of freedom allows better coordination of work, both in individual professional and organization levels. Similar to coordination, this mechanism is enabled by removing or reducing the time-restrictions of professionals' work.

Routine clinic visits can be economically inefficient when the personnel spends time examining patients that do not have any symptoms. With CFUA, the professionals mostly contact patients that have reported significant symptoms. Symptom-based care is connected to the co-creation of health. In co-creation of health, the patient contributes to maintaining or improving their state of health. The co-creation of health mechanism functions also through managing demand.

4. Discussion and Conclusions

This study identified several mechanisms through which CFUA increases value. The mechanisms were analyzed from the healthcare professionals' perspective, and most of the identified mechanisms create value by improving the productivity of the

professionals. The mechanisms affecting efficiency and productivity include both mechanisms improving individual professionals' efficiency in tasks, as well as mechanisms that allow more efficient allocation of tasks between professionals. The implementation and organizational practices seem to have a strong effect on the value creation of digital follow-up applications. The digital application should be an integral part of the professionals' work, with clear roles and procedures for using it. The primary source of value of digital application is the data output that creates or improves information. Incomplete movement of information in an organization deteriorates the value created. Key challenges that digital interventions could solve in healthcare are coordination and integration of fragmented knowledge. In the studied empirical context, CFUA has the potential to improve coordination in multiple ways.

The findings of this study contribute to the development of value-based healthcare and to the evaluation of digital health interventions. The study also has certain limitations. We did not interview the patients. Therefore, the value creation was analyzed focusing on costs rather than health outcomes. However, changes in professionals' work have effects on health outcomes, if the patient work changes. Patient outcomes also affect costs, as quick recovery incurs less costs than prolonged treatment. Furthermore, the actualization of the mechanisms as outcomes or costs was not measured. Nevertheless, this research has identified intermediate outcomes that could be used as measures for the value of digital interventions and for confirming the mechanisms in future research.

References

- [1] M.E. Porter, What is Value in Health Care?, *The New England Journal of Medicine*, **363**(1) (2010), 1–3.
- [2] G. Escrivá Bouley, T. Leroy, C. Bernetière, F. Paquenseguy, O. Desfriches-Doria, M. Préau, Digital health interventions to help living with cancer: A systematic review of participants' engagement and psychosocial effects, *Psycho-Oncology*, **27**(12) (2018), 2677–2686.
- [3] S. Iivanainen, T. Alanko, K. Peltola, T. Konkola, J. Ekström, H. Virtanen, J.P. Koivunen, ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. *Journal of Cancer Research and Clinical Oncology*, **145**(3) (2019), 765–774.
- [4] J.L. van den Brink et al., Impact on quality of life of a telemedicine system supporting head and neck cancer patients: a controlled trial during the postoperative period at home, *Journal of the American Medical Informatics Association : JAMIA*, **14**(2) (2007), 198–205.
- [5] D. Thaker, R. Monypenny, I. Olver, S. Sabesan, Cost savings from a telemedicine model of care in northern Queensland, Australia. *MJA*, **199**(6) (2013), 414–417.
- [6] T. Lizée, et al., Cost-Effectiveness of Web-Based Patient-Reported Outcome Surveillance in Patients With Lung Cancer, *Journal of Thoracic Oncology*, **14**(6) (2019), 1012–1020.
- [7] E. Virkki, *Characteristics of Digital Services Contributing to a Positive Patient Experience in Cancer Care*, Master Thesis, Aalto University, Espoo, 2017.
- [8] WHO, *Classification of Digital Health Interventions v1.0.*, Geneva, 2018. URL: <http://who.int/reproductivehealth/topics/mhealth/en/>.
- [9] R. Pawson, N. Tilley, *Realistic Evaluation*, SAGE, London, 1997.
- [10] D. Denyer, D. Tranfield, J.E. Van Aken, Developing design propositions through research synthesis, *Organization Studies*, **29**(3) (2008), 393–413.
- [11] P. Lillrank, J. Groop, J. Venesmaa, Processes, episodes and events in health service supply chains, *Supply Chain Management*, **16**(3) (2011), 194–201.
- [12] B.G. Glaser, A.L. Strauss, *The discovery of grounded theory: strategies for qualitative research*, Aldine, Chicago, 1967.

Digital Technology Trends Supporting Assisted Independent Living of Ageing Population

Anthony MAEDER¹

College of Nursing & Health Sciences, Flinders University, Adelaide, Australia

Abstract. The paper provides a narrative overview of trends in digital technology related innovations for supporting ageing subjects to live independently with assistance, synthesized from selected scoping reviews and informed by subsequent analysis of peer reviewed literature appearing in the past 10 years. Four categories of trends are identified: assistive and supportive technologies; monitoring devices and systems; communications and connection technologies; and intelligent health information systems. For each of these categories, a synthesis commentary and illustrative examples are provided, concluding with a summary discussion on future directions.

Keywords. Ageing, assistive technologies, digital health, health services

Introduction

Technology for supporting ageing well and care provision for ageing subjects has traditionally been seen as providing additional capabilities to those supplied by human carers [1]. Their potential can be realized through consideration of appropriate alignments between technologies, users and care settings [2]. Three major categories of health purpose for this type of technology-based independent living assistance can be distinguished: enabling personal and social interactions; help with daily living activities; and as a means for delivering clinical services [3]. Recently, use of specific digitally-driven technologies to support ageing subjects in various supported situations has become increasingly adopted [4]. Examples of well-established digital technology for ageing subjects include devices to supplement body functions (e.g. hearing, heart function, diabetes medication), systems for monitoring health conditions (e.g. vital signs, physical activities, mental and sleep states), and services for enabling health and wellbeing interactions (e.g. videoconferencing, online health portals, self-care management apps).

The concept of using technology as a channel for implementing new models of care in ageing has been predominantly based on the notion that it is desirable that the care be provided in non-traditional delivery settings, such as the subject's home [5]. This view leads quite naturally to the interpretation of the role of digital technologies in new models of care in ageing as enabling care delivery to be performed remotely (in location and timing), and with the augmentation of remotely collected information and remote interaction and communication mechanisms [6]. For example, it is recognised that remote patient monitoring offers a sound approach for management of chronic diseases [7], and that remote clinician-patient interactions are suitable for a range of specific clinical tasks [8]. It is also acknowledged that various implementation challenges [9] and adoption barriers [10] exist for these approaches.

¹ Corresponding author: Email: anthony.maeder@flinders.edu.au

1. General Model

There is no apparent standardized nor widely accepted categorization of ageing health-related digital technology areas which can be followed in discussions of associated innovations. Recent literature reviews in the field have tended to aggregate many different application areas rather than conform to a defined structure (e.g. [11], [12]). Attempts to define a universal framework or taxonomy are frustrated by the huge volume and numerous variations in technology types, many having been developed as engineering exercises in non-clinical settings. A recent categorization based on contemporary technology development activities [13] suggested seven clusters: Robotics, General ITC, Sensor technology, Telemedicine, Medication dispenser and Video games. These are somewhat imbalanced for broad scoping purposes, as they vary between highly generic and highly specific topics.

Instead we will align our discussion with four areas of technology usage which are associated with broad types of care purpose, from the perspective of subjects-of-care, and so are distinctive in their orientations of access to or provision of care within the Ageing care ecosystem. These four areas were synthesized initially from a collection of widely-cited scoping reviews and categorization discussions such as those abovementioned. These sources were identified by expert consensus, as an initial step in a more extensive ethnographic analysis of references extracted by structured searching of peer reviewed literature appearing in the past 10 years as reported in detail in [14]. We did not include use of digital technology for business management and non-healthcare operations delivery (e.g. roster scheduling), nor technology used routinely for clinical process support (e.g. patient record systems).

For each of these four areas, we briefly summarize some apparent trends in the recent work related to major use cases, indicative of the overall directions that the field is following. These use cases were chosen due to their high prevalence in literature on ageing and technology and are not necessarily exhaustive, but may be regarded as dominant. We contend that these areas are major “promising directions” in digital technology trends supporting assisted independent living of the ageing population, and can be expected to have a strong impact on the successful implementation of new models of care and health in ageing. The four identified areas are:

- Assistive and supportive technologies which provide physical or cognitive aids to activities undertaken by a subject-of-care as an adjunct component of the activity.
- Monitoring devices and systems which measure and analyze personal health characteristics of a subject.
- Communications/connection technologies allowing a subject-of-care to interact with health carers remotely.
- Intelligent health information systems which empower subjects-of-care to access information and exercise informed control on their health circumstances.

2. Emerging Trends

We next provide commentary on the trends observed for these above four areas, as discerned in our literature survey, including some illustrative examples.

2.1. Assistive and Supportive Technologies

A common need for the elderly is *assistance in mobility* to overcome ambulatory limitations (e.g. wheelchairs). Digital technologies incorporated in mobility devices may be simple such as instrumented canes or frames or may incorporate more sophisticated mechanisms such as motion surfaces, limb supplementation and balance enhancement, depending on the mobility purpose and needs. An important related area is the detection and prevention of falling, which

is often a consequence of compromised mobility and potentially tractable through measurement and prediction. The use of Smartphones for mobility information collection and analysis could offer an easier pathway to delivery. The possibility that routine home care activities could be conducted by an autonomous device such as a robotic assistant has become a popular view, suggesting an appropriate form of generic technology to assist ageing in place.

Decline of *auditory and visual sensory functions* which enable interaction with surroundings cause limitations for older adults, typically addressed through external solutions such as hearing aids. The alternative of internally fitted smart devices such as the cochlear implant is now also established. Progress on augmenting or replicating loss of vision by more sophisticated technology solutions has been slower due to highly challenging neurological mechanisms at the three main sites of intervention: retina, optic nerve and visual cortex. Cognitive decline including loss of memory, balance, location and situational awareness, logical reasoning or understanding of context, through to pathological brain ageing or mental diseases, has been a popular target for technology-based approaches. A range of technology options have been identified, from computer-based sensor-controlled living environments to communication systems enabling involvement of carers in times of need. This approach is sensitive to the specific nature of the cognitive impairment and associated behaviours.

2.2. Monitoring Devices and Systems

Healthy ageing support through digital technology applies measurement and feedback on aspects of healthy lifestyle habits, physiological status, and preventive practices to assist subjects to manage and maintain their condition. This is generally achieved through use of wearable monitoring devices which collect data on variables appropriate to the subject's health circumstances. Other devices can be used to measure physiological quantities or signals which in conjunction with movement information, can allow prediction or detection of adverse circumstances e.g. falling, or cardiac events. A more comprehensive use can be the tracking of performance of conventional daily living activities or the remote observation of individuals by health service providers. Beyond activity tracking, numerous wearable devices for vital signs are also available. It is recognised that there is widespread potential for activity tracking for aged subjects in particular. Future devices may adopt IoT technology enabling greater chance of compatibility and redeployment.

Management of health conditions of individuals (such as chronic diseases) can be conducted outside of clinical environments, using measurement devices located on the subject or in their living spaces which communicate data to a remote clinical decision support system or a clinician. Devices to conduct this telemonitoring function include single and multiple vital signs loggers, heart and brain signal capture, and patient contributed inputs such as alarm buttons. They can also be incorporated into standalone integrated platforms or workstations, which aggregate data locally and provide some limited feedback and analysis to the subject, while communicating and remaining under control of a central clinical agency. The workstation may be used according to a fixed regime (e.g. daily) or when the subject feels it necessary (e.g. if experiencing symptoms). Alternatively multiple devices can be combined in a single portable or wearable system with inbuilt data communication and collection capability, often implemented as body area networks. Incorporation of mobile technologies in this setting offers an easier alternative to integration but is reliant on telecommunications infrastructure. There is argument for a more integrated approach to the future design of such systems to allow for their easier repurposing or extension, for different health conditions.

Ambient Assisted Living (AAL) environments are sensor-equipped and computer-managed living spaces which automatically observe and respond via messages and alerts to health status indicators of their aged and potentially frail or disabled occupants: this concept is also styled as "Health Smart Homes". Target areas for wider application in aged living

include health and environmental monitoring, and providing companionship, social communication and recreation/entertainment. Concerns on ethics of AAL for observing individuals closely and continuously raise issues on privacy and trust, and more generally for user acceptance in health IoT settings. A practical open issue is how large volumes of data generated by such uses can be efficiently and reliably processed by “big data” methodologies.

2.3. Communications/Connection Technologies

In a service environment where person-centred ageing care approaches are becoming more widespread, *telehealth* in various forms offers a natural technology-based mechanism for enabling new model of care solutions. For example, an appraisal of potential expanded usage of telehealth services in UK based on the “Whole Systems Demonstrator” project has suggested that this is a favourable environment for achieving integrated care. Allowing delivery of care by remote clinicians through teleconsultations and teleprocedures is a major component of this area. Another area which has benefited from embracing this approach has been chronic condition management. It has been noted that adoption and acceptance of home telehealth depends on the sensitivity with which the service is delivered, especially regarding the level of support given to subjects and their active involvement. These models also depend on dedicated service elements such as concierge in shared living facilities, or callcentres for broader scale community aged care, and on other appropriate technology solutions which are designed to support users sympathetically with their needs and abilities.

The ability of a computer to substitute for the role of a human is a compelling approach when health tasks are sufficiently simple that human delivery is a waste of resource. A *conversational agent* is a computer program which interacts conversationally (by text or speech) with a subject and is able to satisfy the expectations of the user and the service being delivered: chatbots and avatars are common examples. These have begun to see use in health applications which entail routine human verbal interaction such as self-report observations or medication reminders, with potential in mental health and dementia. Agents for use in satisfying health needs for ageing subjects require particular conversational structures to be effective, and may need to be tuned or adaptive to cater for differences between user preferences and capabilities to ensure effectiveness and acceptability.

2.4. Intelligent Health Information Systems

The trend for *individuals seeking information* about their health condition to make use of Internet sources has accelerated and in general has become an expectation in primary health care where patient-centred activities are preferred. Recently there has been a move to provide websites where information on particular areas of health need (e.g. cancer, diabetes, end-of-life) is aggregated and a compendium of background explanations, guide to clinical; evidence, resources for self-help or group support, and links to external resources are provided. There is a view that the practice by subjects of extending and embedding their knowledge of their health conditions and health histories provides a lifelong mechanism for better engagement, with an implication that ageing individuals who have already developed familiarity with such an approach can benefit from it in their years of declining health. Online sources of information on health conditions can be enhanced with material to support management of conditions and activate patient involvement in their care, and so improve patient outcomes.

Use of technology can provide value in supporting *care delivered in non-clinical settings* which are served by a variety of carers (implied by the last three areas above) in care coordination, telehealth services, and data-driven quality of care monitoring. These aspects of service delivery must be considered in the context of the need for services and the ecosystem which influences their delivery. It has been recognised that using technology to

move the focus of control to the subject can be effective: for example, cooperation by the individual in remote monitoring of chronic disease and behavior change, or in making use of mobile apps to take responsibility for their day to day health care needs.

3. Concluding Discussion

Across the four identified categories there is much commonality in many of the technology components. The use of sensors and wearable devices, and the associated use of smartphones and networked/IoT environments, is a repeated theme. A need exists therefore to ensure that these components can be easily integrated and interconnected, requiring progress in the area of standardization of protocols and data management. Much of the commentary on evolution of new digital technology solutions raises issues related to adoption and compliance, with associated ethical or social imperatives. Development of digital technology solutions in harmony with user expectations and preferences, taking into account limitations on digital literacy and usability for ageing subjects, is essential for creating successful products. While much progress has been made in conventional areas of health informatics such as electronic health record systems, addressing the needs of ageing citizens has generally been less popular. This has been due in part to design challenges across diverse stakeholder groups, and due to adoption reluctance by health services strongly reliant on human-based delivery mechanisms. If digital technology is seen as a means to enable new models of care and improve quality and equity of access, there is potential to accelerate new solution developments in this sector.

References

- [1] Koch S, Hägglund M. Health informatics and the delivery of care to older people. *Maturitas*. 2009;63(3): 195-9.
- [2] Rodeschini G. Gerotechnology: a new kind of care for aging? An analysis of the relationship between older people and technology. *Nursing & Health Sciences*. 2011;13(4):521-8.
- [3] Tinker A. Technology and ageing. In: Stuart-Hamilton I, editor. *An introduction to gerontology*. Cambridge University Press; 2011. p. 363-88.
- [4] Baschke CM, Freddolino PP, Mullen EE. Ageing and technology: a review of the research literature. *British Journal of Social Work*. 2009;39(4):641-56.
- [5] Tinker A, Wright F, McCreadie C, Askham J, Hancock R, Holmans A. *Alternative models of care for older people*. Kings College London: Age Concern Institute of Gerontology; 1999.
- [6] Cheek P, Nikpour L, Nowlin HD. Aging well with smart technology. *Nursing Administration Quarterly*. 2005;29(4):329-38.
- [7] Coye MJ, Haselkorn A, DeMello S. Remote patient management: technology-enabled innovation and evolving business models for chronic disease care. *Health Affairs*. 2009;28(1):126-35.
- [8] Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. *Health Affairs*. 2014;33(2):194-9.
- [9] Karunanithi M. Monitoring technology for the elderly patient. *Expert Review of Medical Devices*. 2007;4(2):267-77.
- [10] Magdalena M, Bujnowska-Fedak UGB. Use of telemedicine-based care for the aging and elderly: promises and pitfalls. *Smart Homecare Technology & Telehealth*. 2015;3:91-105.
- [11] Piau A, Campo E, Rumeau P, Vellas B, Nourhashemi F. Aging society and gerontechnology: a solution for an independent living? *Journal of Nutrition, Health & Aging*. 2014;18(1):97-112.
- [12] Kim KI, Gollamudi SS, Steinhubl S. Digital technology to enable aging in place. *Experimental Gerontology*. 2017;88:25-31.
- [13] Khosravi P, Ghapanchi AH. Investigating the effects of technologies applied to assist seniors. *International Journal of Medical Informatics*. 2016;85(1):17-26.
- [14] Dyer SM, Barnett K, Brown A, D'Angelo S, Johnstone G, Laver K, Lowthian J, Maeder AJ, Meyer C, Moores C, Ogrin R, Parella A, Ross T, Shulver W, van den Berg MEL, Winsall M, Crotty M. Review of innovative models of aged care. Adelaide: Flinders University; 2020 Jan.

Electronic Health Records as a Facilitator for the Implementation of a Strategy for the Reduction of Maternal and Child Morbi-Mortality in Buenos Aires City

Ivan RECALDE^{a1}, Denise LEVI^b, Mariano FRANCO^b, Santiago ESTEBAN^a and Analía BAUM^b

^a*Health Information Management and Statistics Office, Ministry of Health, Buenos Aires City, Argentina*

^b*Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina*

Abstract. This paper describes how an Electronic Health Records (EHR) allowed for the implementation of a strategy to reduce maternal and child morbidity and mortality in outpatient settings of hospitals in the public healthcare system of Buenos Aires City.

Keywords. Electronic Health Records, Prenatal Care, Continuity of Patient Care

1. Introduction

Every day around 830 women die from preventable or treatable causes related to pregnancy or childbirth [1]. In 2016, dealing with unacceptably high maternal mortality indicators, the World Health Organization (WHO) published new recommendations to improve the quality of prenatal care in order to reduce the risk of prenatal deaths and pregnancy complications and increase the probability of a positive outcome for both mother and child. The new model changed maternal and fetal evaluations from four to eight to increase the detection and prevention of complications during pregnancy, childbirth, or puerperium. The identification of biological and psychosocial risk factors allows to provide guidance to pregnant patients on her and her child's health care. This also promotes healthy habits and behaviors by decreasing the probability of prenatal deaths [2].

While in high-income countries the majority of women receive the recommended number of medical visits, 60% of women face the opposite scenario in low-income countries, registering more than half of the world's maternal mortality in sub-Saharan Africa and almost one third in South Asia [1]. Other research warn of the disadvantages of pregnant patients in rural areas in terms of prenatal care and the increased risk of maternal mortality, with their identification to remove barriers to access health systems and guarantee continuity of care being a major challenge [3-5].

1 Ivan Recalde ; E-mail: ivanrecalde@buenosaires.gob.ar .

In Buenos Aires City, 27% of registered pregnancies do not access enough prenatal care appointments and 5% have not had contact with the system until delivery [6]. To improve the number of prenatal care appointments in pregnant patients, the Ministry of Health designed a program called “Camino Verde de la Embarazada”. This was based on the improvement of access to comprehensive care of pregnant patients for an early identification of those conditions or determinants that may endanger the life of the pregnant patients, the fetus, or the newborn. This line of care consists of the recruitment of pregnant patients and in the active search by the primary healthcare (PHC) teams of those patients who are absent on scheduled prenatal care appointments to ensure the performance of the recommended prenatal care.

The program was initially implemented gradually in PHC centers with paper records. It was possible because they work with PHC teams alongside the community. This fact facilitated the early recruitment and monitoring of pregnant patients in its population. Subsequently, the implementation of the program scaled up to the outpatient setting of hospitals. The greatest issue with the implementation in these institutions were the huge amount of patients in comparison to PHC centers, the lack of a process that confirms consultations, and the complications registering them on paper implies. It was difficult to achieve the program, which could only be completed in a hospital, due to the allocation of human resources destined exclusively for this task. The objective of this article is to describe how an EHR allowed a strategy that reduced maternal and child morbi-mortality in outpatient settings of hospitals in the public healthcare system of Buenos Aires City.

2. Methods

2.1. Setting

The public healthcare system of Buenos Aires City is made up of an Ministry of Health, hospitals, and PHC centers structured in program areas in charge of a geo-referenced population. The main objective is to create an integrated network of care, strengthening PHC and more than 200 healthcare teams that provide to the community outside the centers and guarantee access to the system. Since June 2016, an EHR is being gradually implemented in the outpatient setting and as of September 2019, 78 healthcare centers are using EHR with a total of 4,984,680 health records across the health system [7].

2.2. Study Design

This is a descriptive study of how an EHR allowed for the implementation of a strategy to reduce maternal and child morbi-mortality in hospitals of Buenos Aires City. To accomplish this, the project leaders of the Office of Health Information Systems measured pregnant patients' health care processes, flow and registration of information, use of EHR, absenteeism registration in the outpatient offices of four hospitals, by conducting observations and field interviews. With the analysis of the data collected, various alternatives were designed for the registration and output of absenteeism data of pregnant patients on scheduled visits in obstetric outpatient clinics in hospitals, as well as their implementation strategy.

3. Results

The implementation of the “Camino Verde de la Embarazada” program in hospitals included the following aspects:

3.1. *Basic agreement between parties of interest*

Before starting the implementation, a meeting was organized with all those involved: the Ministry of Health team, project leaders of the Office of Information Systems, the heads of external offices, obstetrics office, program areas and PHC centers corresponding to the area as well as the local coordinator of the EHR implementation. The objective was to communicate the guidelines of the program and to agree upon the strategy of its implementation. Then, the schedules related to maternal and child care consultations were obtained, which involved a manual reviewing process in each of the health centers.

3.2. *General Guidelines*

The healthcare network of pregnant patients is organized according to the complexity of each case, with PHC centers being responsible for low-risk prenatal care and directing the attention of high-risk pregnancies to hospitals. The record of patients' absenteeism is carried out by health professionals, using a functionality of the EHR where they can view the list of patients with consultations scheduled and, if necessary, register missed appointments. It also allows them to register gestational age.

3.3. *Automated Report*

An algorithm in the programming language R [8] was developed for the identification of patients who were absent from arranged consultations in the obstetric schedules. Seeking to improve the filter and increase the identification of absent patients, the free text field associated with consultations based on regular expressions was used in addition to the absenteeism functionality, which also recover gestational age. Out of the absent patients identified, the algorithm consumed the Master Patient Index to obtain their basic and contact information. From this, each patient was associated with their corresponding health center under their geographic area, using georeferencing tools.

The report is automatically produced on a weekly basis on a cloud platform mounted on an internal Buenos Aires City government server, which can only be accessed through an authentication process. This protects the confidentiality of sensitive data included in the report. It is composed of the following data: hospital, date, id number, name and surname, date of birth, telephone number, email, address, appointments schedules, gestational age, and corresponding PHC center.

3.4. *Articulation of different levels of attention*

The Ministry of Health team visualizes absenteeism reports from all hospitals. They are responsible for sharing the list of patients and their contact information with program areas. Then, each program area transmits this information to their respective PHC centers. If the patient's address corresponds to other jurisdictions, the team connects the

coordinators of the PHC centers to that corresponding jurisdiction. Once PHC teams receive the information, they will attempt to contact the patient after using gestational age to rank cases, where they can reschedule a consultation in a PHC center or in a hospital according to the complexity of the case.

3.5. Implementation strategy

A pilot test was conducted in the hospital where the program was already working. The solution implemented coexisted with its usual circuit based on paper records, to contrast with the operation of the data output. Once it was found that there was an overlap between both reports, the paper circuit was abandoned and the implementation of the new circuit in other hospitals began gradually (Figure 1).

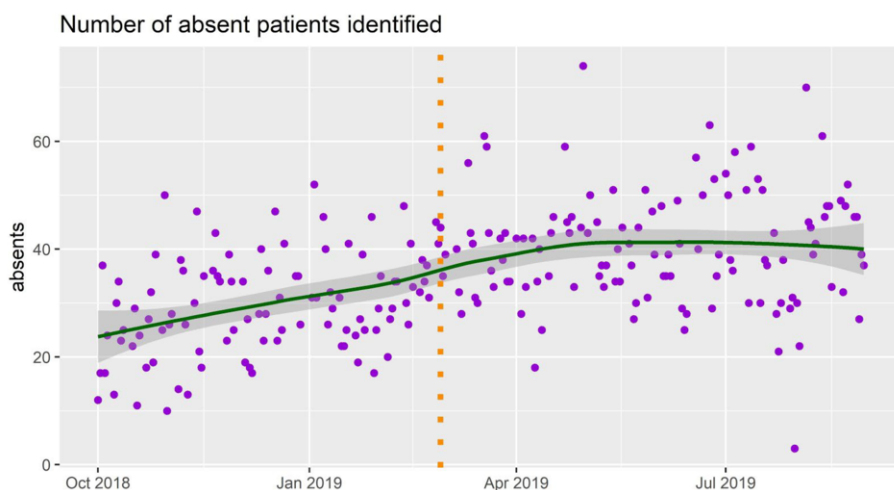


Figure 1. Evolution of the number of daily absent patients identified including nine hospitals of the public healthcare network of Buenos Aires City. The orange dotted line indicates when the new data circuit was implemented in all the hospitals.

4. Discussion

Childhood maternal morbi-mortality rates integrate key indicators of a population's healthcare level and are widely used as measures to represent the effectiveness of a country's health care system. They are constantly used to analyze and monitor the etiologies that accompany, reinforce, reorient, and plan public policies aimed at improving care for mothers and children. It was important to scale the “Camino verde de la embarazada” program to the outpatient setting in hospitals while also conducting low and high risk prenatal care. The main obstacle was to obtain absent patients' data to then carry out an active search.

The EHR played a fundamental role in facilitating the necessary data so that pregnant patients can attend the largest number of prenatal care appointments. Nevertheless, there were several challenges including bringing a standard work procedure to health organizations accustomed to working in a heterogeneous and

independent settings. Another challenge was to bridge different levels of care (outpatient offices in hospitals and PHC teams) with various work modalities.

All this entailed carrying out consensus among multiple parties of interest, making it vital to establish general guidelines and organize initial meetings with everyone involved. In this way it was possible to implement this strategy of active search for pregnant patients in nine hospitals, which is not usual in urban areas. At the same time, an automated report on a weekly basis has the advantage of accessing information faster than before, providing the opportunity for a quick intervention.

The projection this proposal contemplates is that each PHC center can access the report of their corresponding population in real time and a new report with the results of the active search in order to carry out an evaluation of the program. In addition, it would be of interest to investigate the lack of adherence to prenatal care of pregnant patients in Buenos Aires City.

5. Conclusion

The implementation of the EHR was a fundamental tool in achieving an active search strategy for pregnant patients in the hospital outpatient setting with the intention of completing the pre and postnatal care recommended by the WHO in order to reduce maternal and child morbi-mortality. This also allowed the opportunity to manage public policy in nearby jurisdictions, as well as pursue more proactive strategies and not have to wait for patients to appear in the health center as the traditional paradigm had previously defined.

References

- [1] OMS, Mortalidad Materna, 2019 : <https://www.who.int/es/news-room/fact-sheets/detail/maternal-mortality>
- [2] OMS – OPS, Recomendaciones de la OMS sobre atención prenatal para una experiencia positiva del embarazo, 2016 : http://iris.paho.org/xmlui/bitstream/handle/123456789/49550/9789275320334_spa.pdf?ua=1
- [3] Calle Roldán, J, Método de búsqueda activa comunitaria para la captación de gestantes y puerperas en Ecuador, 2017.
- [4] Alizadeth F. , Identifying and Reengaging Patients Lost to Follow-Up in Rural Africa: The “Horizontal” Hospital-Based Approach in Uganda, 2019.
- [5] Hodgins S, Retaining Patients in Care: An Important but Neglected Challenge, 2019.
- [6] ASIS, Gobierno de la Ciudad de Buenos Aires, 2016: http://www.buenosaires.gob.ar/sites/gcaba/files/asis_caba_2016_dic17_vf_1.pdf
- [7] Giussi MV, Baum A, Plazzotta F, Muguerza P, González Bernaldo de Quirós F. , Change management strategies: transforming a difficult implementation into a successful one. *Stud Health Technol Inform.* 2017;245:813-817. PMID: 29295211. DOI: 10.3233/978-1-61499-830-3-813. <http://ebooks.iospress.nl/publication/48264>
- [8] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Evaluating Quality of Electronic Clinical Notes Using the Spanish Version of Physician Documentation Quality Instrument (PDQI-9) in a High-Complexity Hospital. Cross-Sectional Study

Giuliana COLUSSI^{1a}, Amalia BOTTO^a, Romina GIUDICESSI^b, Laura GOMEL^b, Aldo MARANGONE^b, Miguel RODRIGUEZ^b, Sonia BENITEZ^a

^a *Health Informatics Department, Hospital Italiano de Buenos Aires, Argentina*

^b *Medical Auditing Department, Hospital Italiano de Buenos Aires, Argentina*

Abstract. Introduction: There are factors inherent in Health Information Systems that, when are used improperly, can degrade the quality of the information, which may imply, among other things, the lack of integrity, inconsistency or inaccuracy of the information. **Objective:** The aim of this study is to describe the quality of electronic clinical notes using the PDQI-9 score (Spanish version). **Methods:** We evaluated the judgment on the quality of two medical auditors trained in the use of PDQI-9 tool. The unit of analysis was the 3 types of clinical notes: admission notes, progress notes and discharge summaries of adult patients admitted to general wards. Sampling type: random. Statistical analysis: continuous variables were summarized with median and interquartile range. Medians between 2 reviewers were compared using Mann Whitney hypothesis test. **Results:** 120 Electronic Health Record (EHR) were selected, the median of the overall quality of the admission notes were 33 (31-34) and 32 (32-37) for each reviewer respectively, without any statistically significant differences ($p=0.729$). The overall quality of progress notes and discharge summaries were 28 (27-33) vs 27 (27-32), $p = 0.175$ and 33 (32-34) vs 33 (31-34), $p = 0.243$, respectively. **Conclusions:** The quality of the clinical documents evaluated by the trained reviewers using the PDQI-9 score was good. The application of this type of tool is valuable for making a diagnosis in terms of quality that allows improving clinical documentation, as well as peer communication when reading the Electronic Health Record.

Keywords. Electronic Health Record; Documentation; Clinical Note; Quality; Instrument

1. Introduction

The implementation of the Electronic Health Record (EHR), provides benefits regarding the accessibility, availability, and communication of clinical information,

¹ Corresponding author. giuliana.colussi@hospitalitaliano.org.ar

impacting positively on the quality of patient care [1]. However, there are factors inherent in Health Information Systems that, when are used improperly, can degrade the quality of the information [2], which may imply, among other things, the lack of integrity, inconsistency or inaccuracy of the information [3]. An example of this problem is the inappropriate use of copy and paste, as well as the vast amounts of low-quality information automatically inserted in the clinical notes[2].

In this sense, it is essential to audit the quality of clinical records made by health professionals [4], the main role of medical auditors. In our country, we have a validated tool to evaluate the quality of clinical notes in the electronic medical record, but its use is not standardized. This tool, called Physician Documentation Quality Instrument (PDQI-9), evaluates three clinical notes (admission notes, progress notes and discharge summaries), through 9 items using a Likert scale. Using this tool, a higher score indicates a higher quality of the clinical notes[5][6]. The PDQI-9 was adapted and cross-culturally validated in Buenos Aires Italian Hospital so that it can be used in our healthcare context by Internal Medicine specialists [7]. For expanding the use of this tool the aim of this study was to train clinical audits to use this tool in order to evaluate quality of clinical notes.

2. Methods

A cross sectional study was performed from March to May 2018. The Italian Hospital is a center of high complexity located in the City of Buenos Aires, Argentina. It's a school hospital who covers surgical and medical specializations. The informatic development from each one has been made by its own initiative. Since 1998, the Institution designs and builds its own Health Informatic System [8]. EHRs from non-surgical adult inpatients from general wards were reviewed and admission notes, progress notes and discharge summaries were the units of analysis. They were selected by the following criteria:

Inclusion criteria:

- Adult Patients (18 years or more) hospitalized in non surgical general wards with at least 3 days and no more than 15 days from March to May of 2018.

Exclusion criteria:

- No record of admission notes, progress notes and discharge summaries.
- First or last progress note

PDQI-9 consists of 9 items scored using a 5-point Likert scale (1 -not at all- to 5 -extremely-). Responses are summed to create a total score, with higher scores indicating greater quality. The items included are up-to-date, accurate, thorough, useful, organized, comprehensible, succinct, synthesized, and internally consistent.

In our case, considering that 9 is the minimum value that a document can reach and 45 is the maximum, taking into account that these values are the results of a Likert scale, to confer better interpretability and to be able to obtain results as an opportunity for improvement, the team of researchers chose to interpret the results in 5 categories: 9-17 Poor quality, 18-26 Regular quality, 27-35 Good quality, 36-44 Very good quality and 45 Excellent quality.

Two medical auditors received an one-hour training from the research team, once a week for three months. During the training, clinical notes of inpatients were assessed

and then excluded to participate in the analysis of this work. Auditors evaluated the clinical notes applying the PDQI9 SV tool and shared their views about the quality of them. Every instance was moderated by a member of the research team previously in charge of the tool validation process.

Each reviewer individually assessed every clinical note, the corresponding EHR, and filled a Redcap form to complete the assessment.

For statistical analysis, continuous variables were summarized with median and interquartile range. Medians between 2 reviewers were compared using Mann Whitney hypothesis test. The analysis was performed using STATA 15.

The research project was approved by the institutional review board (CEPI #4092). The study was performed in full agreement with current national and international ethical regulations. The researchers declare no conflicts of interest.

3. Results

120 Electronic Health Record (EHR) were selected and reviewed. The median of the overall quality of the admission notes was 33 (31-34) and 32 (32-37) for each reviewer respectively, without any statistically significant differences ($p=0.729$). The overall quality of progress notes and discharge summaries were 28 (27-33) vs 27 (27-32), $p = 0.175$ and 33 (32-34) vs 33 (31-34), $p = 0.243$, respectively. (Table 1)

Table 1. Variation between reviewers about the Global Quality per clinical note ($n = 240$). HIBA, 2019.

	Reviewer 1	Reviewer 2	<i>p value</i>
Admission notes*	33 (31-34)	32 (32-37)	0,729
Progress notes*	28 (27-33)	27 (27-32)	0,175
Discharge summaries*	33 (32-34)	33 (31-34)	0,243

*Median (IIQ25-75%)

Regarding the quality by the domain (Table 2), most of them were around the value 3. Overall, "Organized" was the domain with the lowest value (admission notes, 3.2; progress notes, 2.9; discharge summaries, 3.2). On the other hand, the best-rated domains were "Synthesized" (admission notes and discharge summaries, 3.8), "Internally consistent" (progress notes, 3.3; discharge summaries, 3.8) and "Up-to-date" (discharge summaries, 3.8).

Table 2. Quality by item per clinical note ($n = 240$). HIBA, 2019.

	Admission notes	Progress notes	Discharge summaries
Up-to-date*	3.7 (0.50)	3.1 (0.80)	3.8 (0.43)
Accurate*	3.3 (0.54)	3.1 (0.70)	3.4 (0.51)
Thorough*	3.6 (0.60)	2.9 (0.67)	3.7 (0.50)
Useful*	3.7 (0.50)	3.1 (0.80)	3.7 (0.50)
Organized*	3.2 (0.50)	2.9 (0.64)	3.2 (0.50)
Comprehensible*	3.6 (0.60)	3.1 (0.72)	3.6 (0.50)
Succinct*	3.3 (0.52)	3.0 (0.71)	3.3 (0.50)
Synthesized*	3.8 (0.50)	3.1 (0.82)	3.8 (0.44)
Internally consistent*	3.7 (0.60)	3.3 (0.81)	3.8 (0.50)

*Media (DS)

4. Discussion

The aim of this study was to train medical auditors to use the PDQI-9 Spanish version in order to evaluate quality of clinical notes. Admission notes, progress notes, and discharge summaries from 120 EHRs were reviewed. Most of the time, the items analyzed were around the value 3. When analyzing domains, admission notes and discharge summaries were better scored than progress notes, especially in organization and completeness. This could be explained by the fact that the first two have text entry in structured fields, while progress notes is recorded in free text format. Comparing with other experiences the quality obtained in this study was satisfactory. In results reported by Hahn [9], the highest-rated note attributes were “comprehensible” (4.7) and “accurate” (4.5) and the lowest-rated attributes were “synthesized” (4.2) and “succinct” (4.2). This difference can be due to the training provided to residents in their education program. In this sense, there is evidence that this score is useful as a measure of performance of training programs to improve the quality of clinical record. As an example, in the study carried out at the Children's Hospital of Wisconsin [10] the impact of a training for residents were measured with the PDQI-9 tool. In addition, Miller et al [11], in 2017 published how from a training program they were able to improve the score of the quality above 40 points after training. The objective of the mentioned studies were not only to measure the quality of the records, but also to use the PDQI-9 results as a quality indicator, because a standard score is useful for tracing training strategies to improve the quality of records. Future studies in our setting are necessary to examine the implementation of education programs centered in improve clinical note's quality.

5. Conclusions

The quality of electronic clinical notes evaluated by trained medical auditors using the PDQI-9 SV tool was appropriate. This adapted tool has the potential to improve the documentation using the record quality level as an indicator.

Acknowledgement: We thank to Veronica Peuchot and Sofia Zanetti for their support in this project.

References

- [1] Carnicero J, Andres F. Manual de salud electrónica para directivos de servicios y sistemas de salud. Volumen II: Aplicaciones de las TIC a la atención primaria de salud. CEPAL, Santiago, 2014.
- [2] Baier AW, Snyder DJ, Leahy IC, Patak LS, Brustowicz RM. A Shared Opportunity for Improving Electronic Medical Record Data. *Anesth Analg*. 2017;125: 952–957.
- [3] Embi PJ, Weir C, Efthimiadis EN, Thielke SM, Hedeon AN, Hammond KW. Computerized provider documentation: findings and implications of a multisite study of clinicians and administrators: Table 1.
- [4] *J Am Med Inform Assoc*. 2013;20: 718–726
- [5] Wang A, Dunlop W, Rodda H, Ben-Meir M, Staples M, Walker K. The 9-Item Physician Documentation Quality Instrument (PDQI-9) score is not useful in evaluating EMR (scribe) note quality in Emergency Medicine. *Appl Clin Inform*. 2017;08: 981–993.
- [6] Stetson PD, Morrison FP, Bakken S, Johnson SB, eNote Research Team. Preliminary development of

- the physician documentation quality instrument. *J Am Med Inform Assoc*. 2008;15: 534–54
- [7] Stetson PDB. Assessing Electronic Note Quality using the Physician Documentation Quality Instrument (PDQI-9) *Applied Clinical Informatics*. 2012;3:164–74.
 - [8] Benítez SE. Cross-Cultural Adaptation and Validation of a Score for Evaluating the Quality of Inpatient Clinical Notes [Internet]. Oregon Health & Science University. 2013.
 - [9] F. González, B. De Quirós, D. Luna, A. Baum, F. Plazzotta, C. Otero, S. Benítez, E. Documento de proyecto Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires, (n.d.).
[https://www.hospitalitaliano.org.ar/multimedia/archivos/repositorio/11/recursos/26 TIC en el HIBA](https://www.hospitalitaliano.org.ar/multimedia/archivos/repositorio/11/recursos/26_TIC_en_el_HIBA).
 - [10] Hahn, D., Kolinski, J., Toth, H., Weisgerber, M., Pilon, C., & Wegner, A. Promotion of High Quality Documentation Among Residents on Inpatient Pediatric Wards Using a Standardized Tool to Enhance Faculty Feedback. *Academic Pediatrics*, 2017; 3: 30–30.
 - [11] Hahn, D., Kolinski, J., Toth, H., Weisgerber, M., Pilon, C., & Wegner, A. Delivering Feedback to Residents Using a Documentation Assessment Tool. *Academic Pediatrics*, 2018; 5: 17–18.
 - [12] Miller, R; Fox, T; Mathews, B . Diagnosing our documentation: a novel electronic peer-feedback to improve the quality of hospitalits'notes at a large tertiary care medical center. *Hospital Medicine* 2017,12

Evaluation of openEHR Repositories Regarding Standard Compliance

Maximilian FECHNER^{a,1}, Tobias J. BRIX^a, Tobias HARDT, Johannes OEHM^a,
Martin DUGAS^a and Michael STORCK^a

^a*Institute of Medical Informatics, University of Münster, Germany*

Abstract. We evaluated three repositories implementing the emerging healthcare data standard openEHR for their standard compliance, completeness and vendor lock-in. We found the basic functionality to work consistently across all tested repositories. At the same time, no vendor supports the entire API yet. Some functions like template listing differ slightly in their behavior. Some vendors offer additional custom APIs that are easier to use but lead to vendor lock-in. The openEHR standard itself is designed inconsistently regarding data formats and is missing some basic functionality, for example deletion of templates. With openEHR being a young standard, these issues may be resolved in future releases.

Keywords. Health Information Interoperability, openEHR, Software Validation

1. Introduction

Automated exchange of clinical data between hospitals for improved care and research remains an ongoing challenge. Progress in this area is expected to save billions of dollars via reduced redundancy [1] as well as lead to new medical discoveries e.g. through extensive analysis of the resulting big data sets [2].

The Federal Ministry of Education and Research² in Germany set up the Medical Informatics Initiative³ in 2017 to facilitate data exchange between university hospitals. It consists of several consortia, each of which tries to reach interoperability in a different way. The University of Münster is part of the HiGHmed⁴ [3] consortium that decided to use the openEHR⁵ standard to address this challenge.

OpenEHR is a standard for clinical data storage, retrieval and exchange. It separates clinical modelling for clinicians from IT implementation. Interoperability is achieved by creating common data models. Hospitals that use the same data structures can then directly exchange data without the need for further conversion layers. The countries with the broadest deployments of openEHR are Australia, Slovenia and Norway⁶. Countries pursuing a nationwide eHealth strategy and setting standards for interoperability like Norway seem particularly well suited for the openEHR approach [4].

International cooperation is possible via the “Clinical Knowledge Manager”⁷ (CKM) by Ocean Informatics. HiGHmed uses a governance model designed around a local CKM with clear roles and responsibilities [5] that feeds back parts of its work into

¹ Corresponding Author, Maximilian Fechner, Institute of Medical Informatics, University of Münster, Albert-Schweitzer-Campus 1, 48149 Münster, Germany; E-mail: max.fechner@uni-muenster.de.

² <http://bmbf.de>

³ <https://www.medizininformatik-initiative.de/>

⁴ <https://highmed.org/>

⁵ <https://www.openehr.org/>

⁶ https://www.openehr.org/openehr_in_use/deployed_solutions/

⁷ <https://openehr.org/ckm/>

the international CKM. Other countries like Sweden and Australia also use local CKMs [6].

Several vendors implement the openEHR standard and can in theory be used interchangeably, thus avoiding a vendor lock-in. We evaluated three different openEHR repositories for their standard compliance.

2. Methods

Multiple vendors offer patient repositories based on the openEHR standard in varying degrees of completeness. We selected only systems with support for the openEHR query language AQL, since querying records is crucial for our use-case in HiGHmed: Marand's "ThinkEHR" (now "Better Platform")⁸ (version 2.4.6), Ocean's "LinkedEHR"⁹ (version: 0.9 beta) and DIPS's "EHR-Store"¹⁰ (version: 6.1.0.3166). For the sake of simplicity, these will be referred to by their vendor names, i.e. Marand, Ocean and DIPS.

In an openEHR setting, the following terms are being used. An *EHR* is an Electronic Health Record that acts as a container for a single patient. An EHR contains no patient-data directly, only meta information like a UUID. A *composition* contains a single data record of a patient, e.g. the blood pressure or a microbiology report that is then attached to an EHR. A *template* is the structure definition of a composition. It is the equivalent of a schema definition in other languages, e.g. XSD for XML. Templates are substructured with reusable components called *archetypes* that can be queried independently of the template instances they belong to. To save any given data record into an openEHR repository it is necessary to have an existing EHR-container to attach it to, an existing template definition to validate against, and the data itself in the form of a composition.

We used the official openEHR REST API 1.0.1¹¹ to compare the aforementioned solutions regarding completeness, standard compliance and proprietary functions. The openEHR Repository from DIPS was installed on a local machine. Demo installations of the repositories from Marand and Ocean were offered remotely on the vendor's premises. The following methods of the official openEHR REST API 1.0.1 were tested on all three vendors' repositories in both XML and JSON variants:

- Template: Upload, List, Retrieval
- EHR: Creation with custom ID, Creation with assigned ID, Retrieval
- Composition: Upload, Retrieval, Removal
- Queries: Upload, List, Execution

3. Results

The results of the tests of the different API calls are summarized in Table 1. Green values signify standard compliance. Orange values stand for a non-standard compliant solution. Red means that the function is not implemented yet. Grey values mark functions that are not included in the current version of the openEHR standard.

⁸ <https://platform.better.care/>

⁹ <http://www.oceanhealthsystems.com/products/linkedehr>

¹⁰ <https://www.dips.com>

¹¹ <https://specifications.openehr.org/releases/ITS-REST/latest/index.html>

	Marand	Ocean	DIPS
Template Upload	JSON XML	JSON XML	JSON XML
Template List	JSON XML	JSON XML	JSON XML
Template Retrieval	JSON XML	JSON XML	JSON XML
EHR Creation	JSON XML	JSON XML	JSON XML
EHR Creation with ID	JSON XML	JSON XML	JSON XML
EHR Retrieval	JSON XML	JSON XML	JSON XML
Composition Creation	JSON XML	JSON XML	JSON XML
Composition Retrieval	JSON XML	JSON XML	JSON XML
Composition Removal	Yes	Yes	Custom API
Ad-hoc Query Execution	JSON XML	JSON XML	JSON XML
Stored Query Execution	No	Yes	Custom API
Stored Query Upload	No	Yes	Custom API
Stored Query List	No	Yes	Custom API

Table 1: Results of the openEHR REST API 1.0.1 evaluation

3.1. Templates

Template upload works across vendors in the standard Operational Template (OPT) format (Ocean accepts only archetype_id version > 0). Retrieval of templates is implemented within Marand using the XML representation and DIPS additionally supports the JSON format. Ocean has not yet implemented the template retrieval functionality. The listing of currently available templates works similarly within Marand and Ocean. The returned list contains the latest valid version for each uploaded template and is only available in the JSON format. In contrast, template listing in DIPS returns all versions of all templates including outdated versions. Therefore, the returned list from DIPS has to be filtered manually for currently valid templates to achieve the same result as in Marand and Ocean. The return values differ, too: Ocean returns only “concept” and “template_id”. Marand returns an additional “archetype_id” and “created_timestamp”. DIPS adds a further “id”, “filename”, “type” and “version”.

3.2. Electronic Health Records

EHR creation with automatically assigned ID in JSON format works consistently across all tested vendors. A possible variant using XML format is not documented in the current version of the openEHR standard. EHR creation with custom IDs only works with Marand’s repository. All three vendors support EHR retrieval in JSON format. Marand and DIPS additionally provide retrieval in XML format.

3.3. Compositions

All tested repositories support the creation and retrieval of compositions in JSON format. Besides, Marand and Ocean support XML retrieval. Composition removal in DIPS is only possible using the custom API, whereas Marand and Ocean provide a standard compliant solution.

3.4. Queries

The results of ad-hoc queries show differences in the naming of columns (#0, #1, #2 in Marand versus path names in DIPS and Ocean), but can be performed on all repositories. Standard-compliant stored queries are only possible with Ocean, but DIPS offers a custom solution. Ocean returns only IDs for EHR-queries via AQL, metadata like systemID and creationTime are not supported. The only supported data format for queries in the tested repositories is JSON.

4. Discussion

We successfully evaluated three different openEHR repositories regarding standard compliance and proprietary functions. The examined repositories show differences in their implementation of the official REST API in every evaluated segment (templates, EHRs, compositions, queries). The most important parts of the official openEHR REST API (template upload, EHR creation, composition creation, basic querying) work consistently across vendors. Thus, repositories can be filled with data and data can be queried with the official API calls. At the same time, every tested repository misses some functionality of the official API. At present, Marand does not support stored queries. Ocean cannot create custom EHRs. DIPS cannot remove compositions. Some functions differ slightly in their behavior, such as template listing in DIPS and AQL EHR queries in Ocean. Since openEHR is a comparatively young standard, these issues may be resolved in future software versions.

Additionally, Marand and DIPS both have extensive custom APIs in parallel to the official openEHR API. These custom APIs are in many ways easier to use than the official version, especially regarding compositions. Both vendors offer a custom “flat” composition format. Adoption of this API means that the “load” part of an ETL-process (Extract-Transform-Load) would have to be redesigned upon a change of the underlying repository. Furthermore, DIPS decided to implement some functions exclusively via custom API, like stored queries and composition removal. This approach is not in line with the general openEHR idea of interoperability and may lead to vendor lock-in.

The standard’s API design itself leaves some areas for improvement, too. At present, there are some inconsistencies regarding data formats. Template upload and download are designed around the XML based OPT (**O**perational **T**emplate) format, while template listing returns JSON. Interestingly, DIPS does not even accept its own template JSON output as a valid input. It remains unclear why this design of alternating data formats was chosen. It might derive from merging of different pre-existing APIs.

Template definition also requires a variety of formats. The XML based OET-format (**O**penEHR **T**emplate) is used for the templates themselves and the custom ADL-format^{12,13} (**A**rchetype **D**efinition **L**anguage) describes archetypes. OETs and ADLs can then be combined into the aforementioned OPT, which can then be uploaded into the repositories. The benefits of the creation of the custom ADL format in addition to the mix of JSON and XML that is already in use remain elusive.

¹² https://specifications.openehr.org/releases/AM/latest/ADL1.4.html#_structure

¹³ <https://specifications.openehr.org/releases/LANG/latest/odin.html>

In summary, working with openEHR requires the use of at least three different data formats (XML, JSON, ADL). Harmonization of the used data formats in future versions of openEHR would be beneficial.

Another point regarding API design is the omission of some basic functions. The deletion of EHRs or templates is currently not supported; therefore, unused records and templates can pile up in the systems and might affect maintainability over time.

5. Conclusion

In their current version, the three tested repositories are standard-compliant regarding the essential parts of the openEHR REST API. None of the tested solutions shows complete support of all official API functions.

Marand and DIPS both have extensive custom APIs in parallel to the official API. These custom APIs are in many ways far easier to use than the official version. Furthermore, DIPS decided to implement some functions exclusively via custom API. These proprietary functions might lead to vendor lock-in. In addition, functions to delete templates or EHRs are missing.

Finally, the official API design itself is inconsistent in its use of data formats. At least three different formats (XML, JSON, ADL) are required to work with openEHR in its current form. Since standards like openEHR are in a continuous improvement process, this might change in future releases.

6. Acknowledgements

Supported by BMBF grant No. 01ZZ1802V (HiGHmed/Münster)

References

- [1] J. Walker, E. Pan, D. Johnston, J. Adler-Milstein, D. Bates, B. Middleton, The Value Of Health Care Information Exchange And Interoperability: There is a business case to be made for spending money on a fully standardized nationwide system, *Health Affairs* **24** (2005), W5-10-W5-18.
- [2] T. Murdoch, A. Detsky, The inevitable application of big data to health care, *JAMA* **309** (2013), 1351-1352.
- [3] B. Haarbrandt, B. Schreiweis, S. Rey, U. Sax, S. Scheithauer, O. Rienhoff et al., HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries, *Methods of Information in Medicine* **57** (2018), e66-e81.
- [4] R. Pedersen, C. Granja, L. Ruiz, Implementation of OpenEHR in Combination with Clinical Terminologies: Experiences from Norway, *Int J Adv Life Sciences* **9** (2017), 82-91.
- [5] A. Wulff, B. Haarbrandt, M. Marscholke, Clinical Knowledge Governance Framework for Nationwide Data Infrastructure Projects, *Studies in Health Technology and Informatics* **248** (2018), 196-203.
- [6] G. Ellingsen, B. Christensen, L. Silsand, Developing large-scale electronic patient records conforming to the openEHR architecture, *Procedia Technology* **16** (2014), 1281-1286.

Formal Modelling of FHIR Based, Medical Data Exchange Using Algebraic Petri Nets

Franziska BATHELT ^{a,1}, Michéle KÜMMEL ^a, Sven HELFER ^a, Christian KAMANN ^a and Martin SEDLMAYR ^a

^a*Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden*

Abstract. Enabling interoperability is a challenging task in medical data exchange and is often addressed by the use of versatile communication standards like HL7 FHIR. Although daily routine and scientific experiences show its suitability, the use comes with major risks regarding safety and security of data. To overcome these problems, we present an approach to enable a formal verification of medical communication use cases. We identified the Neonatal Screening as a practical example in which two process participants (physician and screening laboratory) are involved. We analyzed the FHIR specification as well as identified necessary resources in that context and formally modeled them as algebraic specifications. By that, we were able to represent the participants' behavior and data flow with help of Algebraic Petri Nets. This strategy allows to formally verify the correctness of a system by specified requirements regarding data safety and data security.

Keywords. Data Safety, Data Security, FHIR, Algebraic Petri Nets, Neonatal Screening

1. Introduction

Integrated health care depends on both interdisciplinary collaboration of health care providers and communication between their IT systems. Providing the required data interoperability is a crucial task and subject of current research [1].

As an example for a medical process we focus on the screening for Phenylketonuria (PKU), a disorder that is a target of the German Neonatal Screening program [2]. This screening allows the early detection and treatment of congenital metabolic disorders, endocrinopathies and immunodeficiencies in all newborns (given the parents' consent) [3]. For the diagnosis of PKU, a secure and reliable communication between the physician and the screening laboratory is needed. In particular this means, it is necessary to ensure that exchanged data is sent, received and interpreted correctly as well as it is only accessible by authorized participants.

¹Corresponding Author: Email: franziska.bathelt@tu-dresden.de

HL7 FHIR (High Level 7 - Fast Healthcare Interoperability Resources) is a trending important communication standard in health care [4]. It theoretically allows for integrating legal specifications by providing a *Contract Resource* [5], which can include information regarding consent and access restrictions [6].

As these contracts change, IT systems must be manually adjusted. These changes are often tested with the help of simulation methods. However, these methods are usually incomplete, which leads to safety and security problems. This can be countered by formal specifications, which enable the analysis during the concept phase and the possibility to ensure adherence to legal and medical regulations. One possibility to formally specify data exchange processes is to use Algebraic Petri Nets [7] (APNs). APNs simultaneously model control and data flow while being highly comprehensible. It has already been shown that reliable data communication between medical devices can be realized if the devices' internal behavior and data flow are modeled as APNs [8]. We investigate, if it is possible to formally model the participant's behavior and their data exchange for a medical process.

2. Methods

Based on the PKU screening process we proceeded a three step approach.

1. **Modeling of Petri Nets (PN):** The steps of the Neonatal Screening were inferred from the standardized description of the process [3,2] and abstracted for a process model. The participants (physician and screening laboratory) were modeled as Open PN [9].
2. **FHIR Based Data Modeling:** Based on the standardized guideline mentioned above FHIR resources that are necessary for the communication were identified, analyzed and mapped into the structure of algebraic specifications [10].
3. **Modelling of APNs:** The Open PN were enriched with the FHIR-based medical data to model the APNs.

3. Results

3.1. PN Model for Neonatal Screening

The modeling of the Neonatal Screening process, performed by the physician and the screening laboratory, in an Open PN corresponds to the process shown in Figure 4. The physician, who was responsible for the child's birth, has to inform the parents about the Neonatal Screening (p_0, t_0, p_1). The parents can refuse (t_2, p_2) or agree (t_3, p_3). Given consent, the test can be carried out (t_4, p_4). Therefore a sample of blood is transferred to a test card which is sent to the screening laboratory (t_5, p_5). The physician is waiting for the results (p_6). The laboratory physician performs a tandem mass spectrometry (p_{11}, p_{12}, t_{11}) and documents the laboratory experiment at the test card (p_{13}, t_{12}, p_{15}). The result of the examination (p_{14}) can be suspicious (t_{14}, p_{17}) or within normal range (t_{13}, p_{16}). The physician will be informed about every suspicious result immediately by telephone (t_{15}, p_{18}, p_{19}). In both cases, the laboratory physician creates a record of the result (t_{17}, t_{16}, p_{20}) and returns it to the physician responsible ($t_{18}, p_{21}, t_{19}, p_{22}, p_{23}$). When the physician

is informed of the suspicious findings (p_7, t_6), further diagnostic measures are initiated (p_8) to confirm or refute the suspicion. If the findings are inconspicuous, the physician will only be informed in writing (t_{10}, p_{10}) and no further action is required.

3.2. FHIR Based Data for Neonatal Screening

To enable the communication between physicians and the screening laboratory, the following FHIR resources must be considered: (1) The *Patient Resource* includes name and birthdate of the newborn. The Neonatal Screening process also requires communicating the newborns time of birth. As it is not included in the *Patient Resource*, we used the extension *birthTime*. (2) The *ServiceRequest Resource* is needed to inform the laboratory to start the diagnostics. (3) The *DiagnosticReport Resource* includes information about the laboratory results and their interpretation. (4) The *Contract Resource* can include information about access restrictions and consent acceptance.

These resources are connected to further resources; e.g. *DiagnosticReport* might include information about the performer of the diagnostics which can be an organization (*Organization Resource*) or health care provider (*Practitioner Resource*). We abstain from listing all of them in detail for clarity reasons. However, an overview is given in Figure 1.

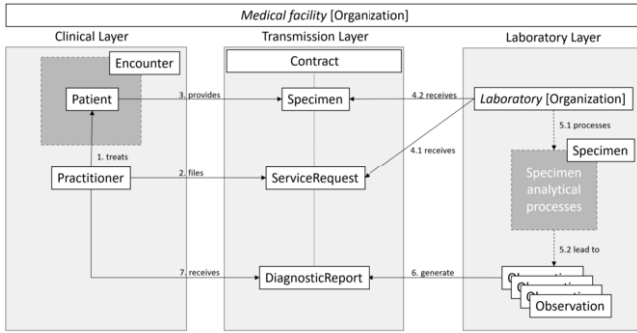


Figure 1. Necessary resources and their connection

By analyzing these FHIR resources, we found a basic structure that can be represented as algebraic specifications, which are shown in Figure 2.

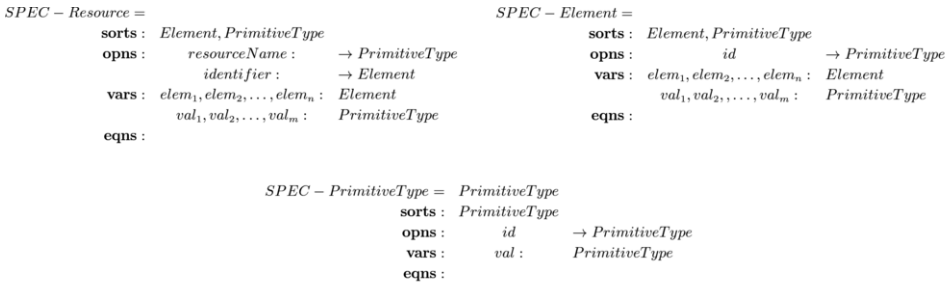


Figure 2. Basic Structure of FHIR resources as algebraic specification

3.3. APN Model for Neonatal Screening

The process of the APN is represented by Figure 4 and the corresponding algebraic specification is described in Figure 3. The algebraic specification is used to formally define data types (e.g. **sorts**: Resource), operations and constants on these data types (e.g. **opns**: $C \rightarrow \text{Resource}$), variables (e.g. **vars**: resourceName: PrimitiveType) and equations (e.g. **eqns**: $\text{resourceName}_C = \text{"Contract"}$). The elements that are defined in the algebraic specification are represented within the PN model, e.g. $\text{arc}(p_1, t_3)$ is labeled with tuple (E, C) , which means that an *Encounter Resource* and a *Contract Resource* are passed.

SPEC – NeonatalScreening =

sorts :	<i>Resource, Element, PrimitiveType, Token</i>	
opns :	<i>physician, lab, E, C, S, O, DR :</i>	$\rightarrow \text{Resource}$
	■ :	$\rightarrow \text{Token}$
vars :	<i>elem_{1,SR}, identifier_E :</i>	<i>Element</i>
	<i>resourceName :</i>	<i>PrimitiveType</i>
eqns :	<i>resourceName_{physician}, resourceName_{lab}</i>	<i>= "Practitioner"</i>
	<i>resourceName_E</i>	<i>= "Encounter"</i>
	<i>resourceName_C</i>	<i>= "Contract"</i>
	<i>resourceName_S</i>	<i>= "Specimen"</i>
	<i>resourceName_O</i>	<i>= "Observation"</i>
	<i>resourceName_{DR}</i>	<i>= "DiagnosticReport"</i>
	<i>elem_{1,SR}</i>	<i>= identifier_E</i>

Figure 3. Algebraic Specification of the Neonatal Screening Process

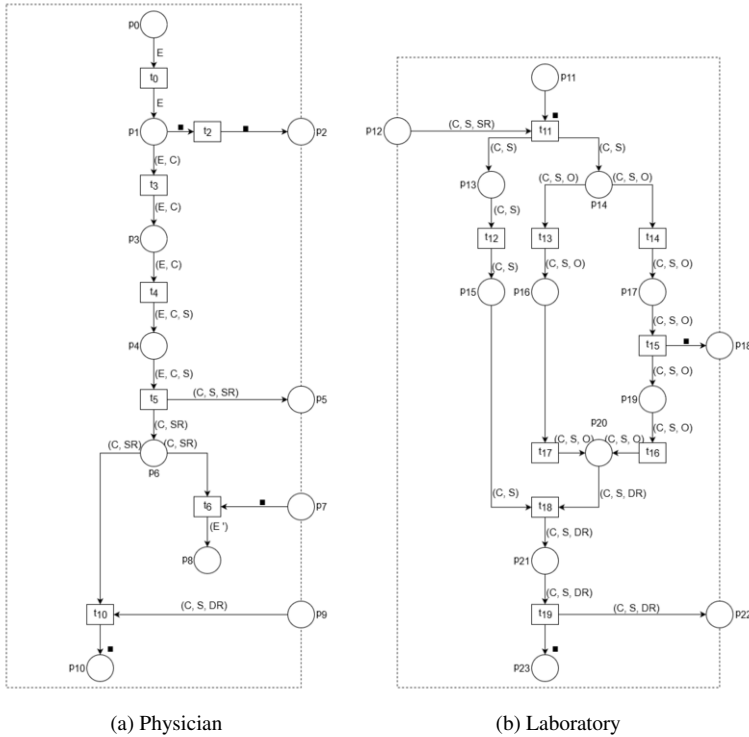


Figure 4. APN for the Physician and Screening Laboratory at the Neonatal Screening

4. Discussion and Conclusion

Based on the Neonatal Screening process for PKU we have shown that a formal modeling of the participants' behavior and their data exchange using APN is possible. By defining a general mapping for FHIR Resources into algebraic specifications a formal data treatment within the concept phase of a system that communicates with FHIR-conform messages becomes possible. The use of the *Contract Resource* enables the inclusion of safety and security requirements into the initial specification. This forms the base to ensure a safe, secure and reliable communication.

The example of PKU screening we described is somewhat limited: It focuses on in-patient births with the test taking place within the first 36-72 hours after birth and we did not include other disorders (e.g. atypical forms of PKU) [3]. Also the modeling process itself leads to an abstraction of single steps. We performed these abstractions in agreement with medical experts.

Albeit these limitations, our approach is a necessary step to formally treat data and its exchange. This will allow formal verification of health care systems. There are currently no approaches but one on data formalization within the composition of communicating systems [8]. To further improve our approach, we aim to include possibilities for representing and verifying time constraints and allow for a real-time adaption process to handle atypical courses of a disease. This could make it possible to test if the prerequisites of a clinical guideline can be met (e.g. process participants, time limits, technology). Also the effects of changes in the guidelines can be theoretically verified and therefore adjustments can be made (e.g. other process participants, changed process steps). These validated guidelines could not only improve the efficiency of health care systems but also patient safety.

References

- [1] Prokosch HU, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine: A Large Data Sharing Network to Enhance Translational Research and Medical Care. *Methods of Information in Medicine*. 2018;57:e82–e91.
- [2] van Wegberg AMJ, MacDonald A, Ahring K, Bélanger-Quintana A, Blau N, Bosch AM, et al. The complete European guidelines on phenylketonuria: diagnosis and treatment. *Orphanet Journal of Rare Diseases*. 2017;12(1):1–56.
- [3] Nennstiel-Ratzel U, Genzel-Boroviczeny O, Böhles H, Fusch C, Grüters-Kieslich A, Mohnike K, et al. Neugeborenen-Screening auf angeborene Stoffwechselstörungen und Endokrinopathien - Leitlinie 024/012. AWMF online; 2011. Available from: https://www.gnpi.de/leitlinien/aktuell/024-012L_S2k_Neugeborenencreening_2011-12.pdf.
- [4] Lehne M, Luijten S, Vom Felde Genannt Imbusch P, Thun S. The Use of FHIR in Digital Health – A Review of the Scientific Literature. *Studies in Health Technology and Informatics*. 2019;p. 52–58.
- [5] HL7 org. FHIR Contract; 2019. Available from: <https://www.hl7.org/fhir/contract.html>.
- [6] Pfiffner PB, Pinyol I, Natter MD, Mandl KD. C3-PRO: Connecting ResearchKit to the Health System Using i2b2 and FHIR. *PLOS ONE*. 2016;11(3):1–8.
- [7] Reisig W. Petri nets and algebraic specifications. *Theoretical Computer Science*. 1991;80:1–34.
- [8] Bathelt-Tok F, Gruhn H, Glesner S, Blankenstein O. Towards the Development of Smart and Reliable Health Assistance Networks Exemplified by an Apnea Detection System. 2014 IEEE International Conference on Healthcare Informatics – Proceedings; 15-17 September 2014, Verona, Italy.; 2014.
- [9] Reisig W. Understanding Petri Nets: Modeling Techniques, Analysis Methods, Case Studies. Springer; 2013.
- [10] Ehrig H, Mahr B. Fundamentals of algebraic specification 1: Equations and initial semantics. vol. 6. Springer Science & Business Media; 2012.

Health Metrics Network (HMN-WHO), a Tool to Assess the Quality of the Information System in Burkina Hospitals

Seydou Golo BARRO^{a,c,1}, Amadou DICKO^a, Paul-André DE LAMÉ^b and Pascal STACCINI^c

^aNazi Boni University, Bobo-Dioulasso, Burkina Faso

^bTraining and Research Center of Medical Technology, Bobo-Dioulasso, Burkina Faso

^cRETINES Lab - IRIS department - Medical Faculty of Nice University, France

Abstract An assessment of the quality of the information system was conducted in the 14 hospitals of Burkina Faso in 2017 using the HMN tool (Health Metrics Network). This evaluation was part of the process of developing a master plan for implementing a HIS for the hospitals in the country, and was aimed at analyzing the sub-components of the existing information system. The results are presented as scores, one per component, evaluated in % and converted into quartiles, ranging from the 1st quartile "Not adequate at all" to the 4th quartile "Highly adequate". The scores are as follows: Resources = Q1 (29%); indicator = Q2 (46%); data sources = Q1 (28%); information production = Q1 (21%); data management = Q2 (47%); dissemination and use = Q3 (55%). In conclusion, overall the system is not adequate (mean score 39%), particularly at the level of information production (score 21%).

Keywords: HMN Tool, Information System Quality, Hospitals, Burkina Faso

1. Introduction

In the health sector, it is critical to have high-quality data on performance available routinely [1-3]. Improving the performance of a country's national health information system depends on the quality of its healthcare information system. To address this challenge, the Ministry of Health of Burkina Faso recommended that each Hospital Center develops a Master Plan for implementing their Hospital Information System (MPHIS) [4,5]. It is within this framework that we conducted a study to evaluate the quality of the information system in terms of data collection and management in the fourteen (14) hospitals the country counts. The aim of this study was to analyze the current data collection and management processes to have a baseline evaluation, and then to follow its progression. In order to measure the performance of the system's components and ensure the comparability of the results between countries, the World Health Organization (WHO) and its partners have created an evaluation tool called "HMN" (Health Metrics Network). It has two parts, the normative portion (components and standards) and an implementation portion (a roadmap) [5]. The purpose of this tool is to create a harmonized framework for the development of a national health

¹ Corresponding Author, Email: golo-seydou.barro@univ-amu.fr.

information system that meets specific standards. The HMN methodology provides a comprehensive assessment of the system's environment and organization, the responsibilities, roles and relationships between actors as well as the technical and technological aspects related to data collection and analysis. The organizational, functional and technical components are the elements that we evaluated.

2. Methods

The evaluation focused on the following six sub-components: (i) background and resources: policy, regulatory and financing aspects as well as the available infrastructure and resources; (ii) indicators: basic indicators of disease management and health promotion; (iii) data sources: population census, vital statistics, population surveys, patient records and registers, care units reports; (iv) data management: data collection, sharing, and storage; (v) information production: quality of the data, especially aspects of transparency, disaggregation, representativeness, consistency, periodicity, timeliness, completeness and collection method; (vi) dissemination and use: information dissemination flow, use for decision-making, including policy development, advocacy, planning, implementation of interventions as well as follow-up and evaluation. Regarding data collection of the relevant information, different forms have been designed. The data collection was carried out by groups of people assessing a specific component of the tool in each hospital. The number of assessors varied from 3 to 5 depending on the importance of the component. People in each group were asked to assign a score between 0 and 3 to each item of the component they were assessing and then to compute the average score per item. Collected data were recorded in the electronic version of the tool programmed to provide the results of the evaluation automatically. The results were converted into four groups or "quartiles" defined on an ordinal scale and reflecting the evaluation of the performance of the system by the respondents: group 1, score $\leq 25\%$, "Not adequate at all", corresponds to the first quartile; group 2, between 26% and 50% "Present but not adequate", corresponds to the second quartile; group 3 "Adequate", score between 51% and 75% representing the third quartile; and group 4 "Highly adequate", score between 76% and 100%, corresponding to the fourth quartile. The main difficulties encountered in our study were the unavailability of respondents, access to information and documentary evidence. The results of the evaluations are presented as bar charts and tables.

3. Results

The results of the evaluation showed some achievements, but also many shortcomings. For better consistency, the results of some components have been combined to make it easier to understand.

3.1. Components resources, indicators and data sources

The evaluation showed that both resources (32%) and indicators (46%) were not adequate with a very low score for implemented policies (16%), Table 1.

Overall, the data sources are not adequate (38%), Table 2. The sources that are adequate are those for which activities are systematically performed. These are the inventory of resources (55%). Civil status is the least adequate source (10%).

Table 1: Resources and indicators component

Categories	Max i	Score	%
Policy and Planning	30	4,9	16
HIS institutions, human resources and financing	15	8,8	59
HIS Infrastructure	12	4,4	37
Total	57	18,1	32
Indicators	15	6,9	46

Table 2: data source component scores

Data source	%
Census	32
Vital statistics	10
Population-based surveys	41
Health and disease records	37
Health service records	50
Resource records	55
Total	38

3.2. Dissemination and use of data component

Overall, use and dissemination are adequate (55%), Table 3. This result is mainly attributable to the use of information for action and advocacy (64%) as well as to the use of data analysis tools (61%). Use of information for resource allocation as well as use of information for action is low (44% and 45%). These unsatisfactory results may be explained by the fact that decision makers require complete, timely, accurate, useful and validated information, which is not always the case in hospitals. The integrated HIS analytical reports are produced and cover most of the main health indicators.

Table 3: Presentation of the dissemination and use of information data component

Categories	Maxi	Score	%
Analysis and use of information	12	7,3	61
Information use for policy and advocacy	3	1,9	64
Information use for planning and priority setting	6	3,5	59
Information use for resource allocation	6	2,7	45
Information use for implementation and action	12	5,2	44
TOTAL	39	20,7	55

3.3. Information products component (quality of information)

The mean score of this component is 21%.This score, by the tool standards is considered as "Not adequate". Only "periodicity", "representativeness" and "data-collection methods" criteria fall in the adequate zone. The other criteria are not adequate and are in the second quartile. These are: promptness, completeness, and method of estimating quartile information.

3.4. Evaluation summary

The evaluation summary results presented in Figure 1 shows that the quality of the information system is not adequate. The overall mean score is in the 2nd quartile (39%), so "Not adequate". The best score of the system is related to dissemination and use of information. The worse score is related to information production (21%).

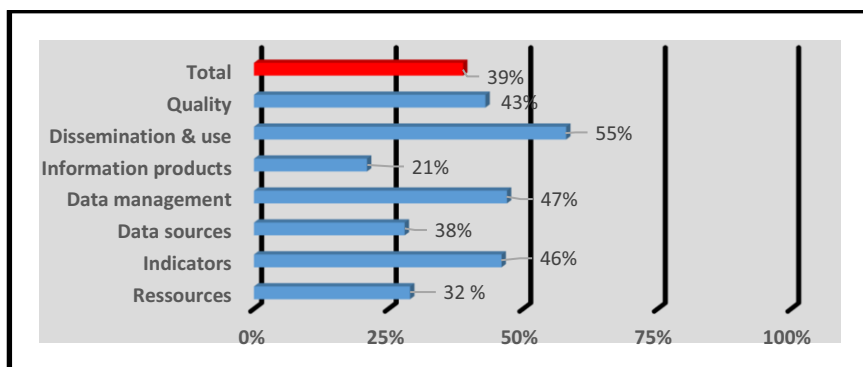


Figure 1. Overall information system indicators scores

4. Discussion

The National Health Information System of Burkina Faso is not efficient, particularly regarding its hospital sub-system. This was the rationale for this evaluation, which allowed to identify its strengths and weaknesses. The strengths of the hospital information system are a good organization of the services, a legislation related to mandatory reportable diseases, tools for collecting information, databases built with DHIS2 [4,5], procedures for implementing the principles of official statistics, effective means for cross consultation at all levels. There are also weaknesses, for example not including some parameters such as civil status. Social insurance is very limited. Moreover, there is no systematic evaluation of the hospital information system contrary to what is done in the peripheral health structures. Indicators of health status (maternal and infant mortality) are not adequate at all (21%). This is partly due to the lack of performance of the civil registration process and the lack of a good certification system of the causes of death [6]. In addition, administrative statistics do not always guarantee the quality of certain indicators such as deliveries attended by qualified personnel. There are indicators discrepancies revealed during the surveys [7]. It should also be noted that the results evidenced the existence of a true coordination problem in the management of information as reported in Botswana in 2015 [8] and in the study by Lippeveld T et al. [9]. There are also problems due to the lack of equipment, financial resources and qualified personnel as reported by Verberk et al in 2004 [10]. Of the 14 hospitals, Bogodogo District Hospital and Tengandogo District Hospital have the highest scores with 62% and 69% respectively. We must mention that these two hospitals are recent institutions that benefitted from the support of certain partners such as the Chinese cooperation. The indicators of data dissemination and use got the highest score in our evaluation (58%). However, these scores remain low by current standards and show an underutilization of information in the hospitals of Burkina [11]. The results of the evaluation show the total score of the components is 39%. A similar score was found in an evaluation carried out at the University Hospital Center of Oran in Algeria (37%) [12].

5. Conclusion

The evaluation of the hospital information system using the HMN tool has identified indicators measuring the performance of the information management system against current standards and practices. The analysis of the existing system made it possible to identify its strengths and weaknesses to extrapolate the needs for the design of the future system. It also evidenced the achievements and efforts that remain to be made in order to have an efficient and effective information system. The evaluation showed the weakness of the activity coordination, the poor quality of certain indicators, the insufficient exploitation of the databases and the low level of data production and dissemination of information. These are all the reasons that justify the development and implementation of a Master Plan for a Hospital Information System (MPHIS). This will have the advantage of unifying the efforts, coordinating the actions and harmonizing the information management at the national level. The implementation of the MPHIS also has the advantage of harmonizing the population numbers (reference population), to strength the quality of the subsystems and to improve dissemination of information. Health data sciences and mobile health (mhealth), will be an important contribution to the management of massive data produced in hospitals [10,11,13].

References

- [1] [WHO, *Data Quality Review*, September 2017.
- [2] J. Alipour and M. Ahmadi, Dimensions and assessment methods of data quality in health information systems, *Acta Medica Mediterr* 33 (2017), 313–320.
- [3] J. Lungo & F. Igira, Development of health information system in Zanzibar: practical implications. *Journal of Health Informatics in Developing Countries*, 2 (2008).
- [4] Direction générale des études et des statistiques sectorielles, Annuaire statistique 2017 du Burkina Faso, Ministère de la santé, 2018.
- [5] Burkina_Faso_National_Health_Strategy_2011-2020_French.pdf
- [6] S. G. Barro, J. C. Dufour, M. Somda, S. Malo & P. Staccini, Cause-of-death certification in Africa: analysis of a model implemented at the teaching hospital of Sourou Sanou of Bobo-Dioulasso (Burkina Faso), *Stud. Health Technol. Inform* 205 (2014), 1110-1114.
- [7] Enquêtes et recensements. <http://www.insd.bf/n/index.php/publications?id=83>.
- [8] O. Seitio-Kgokgwe, R. D. C. Gauld, P. C. Hill & P. Barnett, Development of the National Health Information Systems in Botswana: Pitfalls, prospects and lessons, *Online J. Public Health Inform* 7 (2015), e210.
- [9] T. Lippeveld, R. Sauerborn, C. Bodart, Design and implementation of health information systems <http://apps.who.int/iris/handle/10665/42289>.
- [10] F. Verbeke, P. Shamashanga, C. Amisi, G. Karara, Developing a national e-health strategy for DR Congo: a preliminary analysis of business needs, existing information systems and solutions, *Journal of Health Informatics in Africa* 2 (2014).
- [11] G. S. Barro, Certification des causes de décès en Afrique: Analyse de modèle au CHU Sourou Sanou de Bobo Dioulasso, Burkina Faso, Aix-Marseille, 2014.
- [12] S. Chougrani, S. Ouhadj & F. Agag, Évaluation du système d'information hospitalier de l'Établissement hospitalier universitaire d'Oran, *Sante Publique (Bucur.)* 25 (2013), 617-626.
- [13] D. Blumenthal, Stimulating the adoption of health information technology, *N. Engl. J. Med* 360 (2009), 1477-1479.

Implementation and Analysis of Two Knowledge Base Approaches for the Treatment of Chronic Wounds

Stefan VOGEL^{a,1}, Jendrik RICHTER^a, Stefanie WACHE^a, Kerstin PISCHEK-KOCH^a, Simon AUCHTER^b, Sebastian ZEBBITIES^b, Karen GÜTTLER^b, Ursula HÜBNER^c, Mareike PRYZSUCHA^c, Jens HÜSERS^c, Björn SELLEMANN^d

^aDepartment of Medical Informatics, Univ. Medical Center Göttingen, Germany

^bAtacama blooms GmbH & Co. KG, Germany

^cHealth Informatics Research Group, University AS Osnabrück, Germany

^dMünster School of Health, Univ. of Appl. Sciences Münster, Germany

Abstract. The access to data in healthcare is an enabler for the implementation of clinical decision support systems (CDSS) in practice. The usage of CDSS aims to be of efficient assistance to healthcare providers. The aim of the BMBF project “PosiThera”, is to support the involved professions in the treatment process of chronic wounds. In this study we implemented the formalized knowledge of chronic wound diagnosis into two different knowledge base approaches, the HL7 Arden Syntax and a Petri net approach. The motivating factor behind our study was to use both approaches for the implementation of the projects knowledge base and to compare the results. We implemented the formalized knowledge successfully in both approaches. The results of our comparison showed similarities and differences of the Arden Syntax and the Petri net approach, which might support the evolution of both approaches in the future.

Keywords. Clinical Decision Support, Arden Syntax, Petri net, Chronic Wound

1. Introduction

IT infrastructures have become increasingly important factors in the healthcare system especially for information exchange. The changing paradigm from the paper-based health record to the electronic health record promotes the rising of clinical decision support systems in practice, as the access to data is a crucial requirement for the processing of decision support algorithms. Key components of clinical decision support systems are the knowledge base and the inference engine. Both are required to implement and process knowledge for a CDS System. [1, 2, 3]

The BMBF project “PosiThera aims to develop a CDS System for the diagnosis and treatment of chronic wounds [4]. The research aim of this study was to develop and compare two different implementations of the project’s knowledge base and inference engine by using HL7 Arden Syntax and a Petri net approach.

¹ Corresponding University Medical Center Göttingen, Department of Medical Informatics, D-37075 Göttingen, Germany; E-mail: stefan.vogel@med.uni-goettingen.de.

2. Methods

The development process of our system was based on the knowledge engineering process recommended by Waterman [5] and Durkin [6], which is structured into the steps: (1) Problem assessment, (2) Data and knowledge acquisition, (3) Development of a prototype system, (4) Development of a complete system, (5) Evaluation and revision of the system, (6) Integration and maintenance of the system.

This study gives some results for step 3 for the use case diagnosis of chronic wounds identified in step 1 [7]. The knowledge implemented in the prototype was obtained by analyzing S3 guidelines (3), publications (5), and literature (3) in the context of chronic wound treatment to extract and formalize the knowledge with the focus on our use cases (Step 2). In the formalization process we developed 60 production rules and 3 decision trees.

To implement the formalized knowledge we tested two approaches and compared their performance based on a list of project specific criteria, which are as follows: Binding of Data Ressources, Logical Operators, Medical Logic Organization, Human Interpretable, Knowledge Base Scalability, Maintainability, FHIR Connection, Forward Chaining Algorithm implementable, Backward Chaining Algorithm implementable, and Uncertainty (Fuzzy Logic).

HL7 Arden Syntax was selected, because it is an established and standardized approach to implement machine executable medical knowledge [8, 9]. The Petri net approach was chosen, because it is already implemented in a software product of Atacama Software GmbH in the context of nursing.

2.1. Approach HL7 Arden Syntax

Arden Syntax is a HL7 standard for computerized knowledge representation and processing [8, 9]. It can be used to create knowledge bases for CDS systems and enables the execution of machine-readable medical logic modules (MLM). The main components of every MLM are Maintenance, Library, Knowledge, and Resources.

“Maintenance” contains metadata about the MLM, like version identifier, first implementation, and update. “Library” contains information about the MLM’s context. Crucial for example is in which use case the MLM should be used. This requires knowledge sources and institution specific source linkages used to implement them. The “Knowledge” section represents the processing logic of the MLM, which is expressed as production rules. “Resources” allow to define different language specific messages.

Within this study we concentrate our efforts on the knowledge section.

2.2. Approach Petri net

Within the work of a Bachelor thesis at the University of Bremen in cooperation with the company Atacama Software GmbH [10], an implementation of an extended Petri net and a programming language have been developed, which are now also used within the product apenio by Atacama Software GmbH. The extended Petri net enables defining and processing logical rules modelled after most knowledge bases. Each net is defined by: Places, Arcs (flow relations) and Transitions.

Any place is being mapped with a logical entity of the defined logical structure and can be marked with one or more tokens. A token represents a unique object of a place and can have varying attributes. Arcs are directed flow relations between a transition and

a place. They can contain Boolean expressions, which are executed on the tokens of their place. The place with an arc aiming on a transition is called input, otherwise output place. If the conditions of a transition are fulfilled, then it is called activated. An activated transition can fire. Firing means that tokens of the input places get consumed and a token in the output place will be generated. Based on a starting state of a Petri net with multiple tokens in some places (called marking) there is the possibility, that one or more transitions in the net can be fired and a new marking can be reached as a result. An example for the firing process of the petri net is illustrated in figure 1.

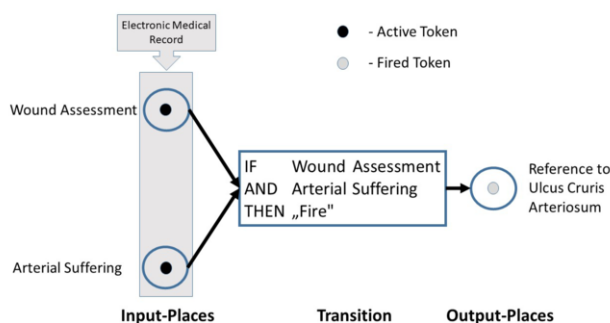


Figure 1. Firing Process of the Petri net approach

3. Results

We were able to implement the knowledge base on the use cases diagnosis and therapy of chronic wounds successfully in both approaches. Based on our criteria list we compared the Petri net approach of Atacma Software GmbH and HL7 Arden Syntax. The results of the comparison are summarized in table 1.

3.1. Similarities

In our study we could identify a variety of similarities between the Petri net and Arden Syntax approach. Medical Logic Modules build up the core to implement medical knowledge in the Arden Syntax. These logic modules are highly standardized (see chapter 2.1). In our use cases it was necessary to organize the modules in file archives which are related by usage context. For example, an archive would consist of causal diseases of an Ulcus Cruris Arteriosum or Ulcus Cruris Venosum. In comparison the Petri net approach is not as highly standardized as the Arden Syntax, but the logic modules can be organized as well in an archive hierarchy. This organization in both approaches enables a sufficient maintainability and scalability. Both approaches use a versioning of Logic Modules to support the developer in cases of maintaining the knowledge base.

The Arden Syntax as well as the Petri net approach operationalize the logic modules by production rule algorithms. These can be implemented as forward and backward chaining procedures, which make both approaches quite flexible to solve complex decision problems.

One benefit of the Arden Syntax is the high performance in human interpretability. The Petri net approach can be supported by easy to read excel lists which contain the medical logic. This approach is different in comparison to the Arden Syntax, but also enables non-informatical professions to understand and discuss about the module contents.

For our implemented use cases both toolboxes of logical operators were sufficiently fitting.

3.2. Differences

While there are a lot of similarities, there are as well differences of the approaches. We could identify, that the binding of data resources is handled differently by both approaches. The Arden Syntax binds data resources separately in every medical logic module by a curly braces statement, which refers to a data source. The Petri net of Atacama Software GmbH uses a central data binding module. The benefit of a central data binding module is to avoid redundancies and could lead to higher transparency in more complex knowledge bases.

Decision scenarios under uncertainty can be solved through the Arden Syntax since version 2.9 by the implementation of fuzzy logic. The Petri net solution does not implement fuzzy logic but can be extended by a Bayes net to deal with decisions under uncertainty.

The usage of the HL7 standard FHIR is implemented in at least one commercial work suite solution of the Arden Syntax as a FHIR connector. The implementation of communication standards in context of scalable data-exchange is beneficial in cases of clinical decision support applications, for example by connecting the CDS System to a hospital information system (which can provide a FHIR interface).

The Petri net approach does not provide a FHIR interface, but a REST interface.

Table 1. Comparison Table of the Approaches Petri Net and Arden Syntax

Criteria	Petri net	Arden-Syntax
Binding of Data Ressources	Central Data Binding Module	Curly Braces in every MLM
Logical Operators	Fitting for the use cases	Fitting for the use cases
Medical Logic Organization	Blocked Production Rules, Archive	Medical Logic Modules, Archive
Human Interpretable	Yes	Yes
Knowledge Scalability	Yes, but increasing conflict management of rules	Yes, but increasing conflict management of rules
Maintainability	Logic Modules Versioning	MLM Versioning
FHIR Connection	Customizable	FHIR Connector in commercial product,
Forward Chaining Algorithm implementable	Yes	Yes
Backward Chaining Algorithm implementable	Yes	Yes
Uncertainty (Fuzzy Logic)	Customizable	Yes

In conclusion, we could identify that the Petri net as well as the Arden Syntax approach are quite similar in their functionality. The main differences we were able to find in the areas FHIR Connection, Binding of Data Resources and Uncertainty.

4. Discussion

In the context of our project PosiThera the study was quite helpful for the future work relating to the evolution of our system with focus on extending the knowledge base. The results show gaps of both approaches as well as benefits and desired functional objectives, which may become relevant for further improvement in the future. Especially the Arden Syntax as established standard could benefit from the results of our study. Suggestions to evolve the Arden Syntax could be to provide a standardized FHIR interface and develop a more fitting solution for the binding of data resources. A second suggestion might also be a good start to overthink and advance the semantic interoperability issue of the Arden Syntax, known as “curly braces problem” [11].

5. Conclusion

In this study we were able to show similarities and differences of the approaches HL7 Arden Syntax and the Petri net approach of Atacama Software GmbH. Both were used to implement our projects use cases. For the ongoing project we decided to use the Petri net approach to benefit from existing interfaces like the connection to a terminology server. What we discern from our study is, that the knowledge modules could be easily transferred from the Petri net approach to the Arden Syntax as well as the opposite direction, which might be of benefit for our future work.

Acknowledgements

This study was funded by the Federal Ministry of Education and Research Germany, support program: Medizintechnische Lösungen für die digitale Gesundheits-versorgung (grant: 13GW0171C).

References

- [1] R. Greenes, *Clinical Decision Support: The Road to Broad Adoption*, 2nd edition, Academic Press, Amsterdam, 2014.
- [2] P. Roshanov, N. Fernandes, J. Wilczynski, B. Hemens, J. You, S. Handler et al., Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials, *BMJ* **346** (2013), f657.
- [3] B. Chaudhry, J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth et al., Systematic review: impact of health information technology on quality, efficiency, and costs of medical care, *Ann Intern Med* **144** (2006), 742-752.
- [4] U. Hübner, M. Przysucha, S. Vogel, J. Hüsers, S. Wache, K. Güttler et al., Intelligente Versorgung von Menschen mit chronischen Wunden, *mdi – Forum der Medizin_Dokumentation und Medizin_Informatik* **19** (2017), 108-111.
- [5] J. Durkin, *Expert Systems Design and Development*, Prentice Hall, Englewood Cliffs, NJ., 1994.
- [6] D. Waterman, *A Guide to Expert Systems*, Addison-Wesley, Reading, MA, 1986
- [7] S. Vogel, M. Przysucha, S. Wache, J. Hüsers, U. Hübner, B. Sellemann, Decision Support Use Cases for the Treatment of Patients with Chronic Wounds, (2018), doi: 10.3205/18gmds025.
- [8] R. Jenders, K. Adlassnig, K. Fehre, P. Haug, Evolution of the Arden Syntax: Key technical issues from the standards development organization perspective, *Artificial Intelligence in Medicine* **92** (2016), 10-14.
- [9] Arden Syntax v2.10 (Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10).

- [10] S. Auchter, Bachelorthesis: Erfassung und Repräsentation von Pflegekomplexmaßnahmen-Scores (PKMS) auf Basis von Petrinetzen, University Bremen – Faculty 3 Mathematics and Computer Science, 2018.
- [11] M. Samwald, K. Fehre, J. de Bruin J, K. Adlassnig, The Arden Syntax standard for clinical decision support: Experiences and directions, *Journal of Biomedical Informatics* **45** (2012), 711-718.

Information Capturing in Pre-Hospital Emergency Medical Settings (EMS)

Kerstin DENECKE^{a,1}, Lea MEIER^a, Jan G. BAUER^a,
Matthias BENDER^a, Christopher LUEG^a

^a Bern University of Applied Sciences, Bern Switzerland

Abstract. Emergency medical situations are characterized by high physical, cognitive and mental demands on the paramedics on the ground. Studies suggest that crucial information such as treatments administered to patients is often documented retrospectively, during patient transport or once a patient is handed over to an emergency department. Information access may also be surprisingly difficult (e.g. patient medical history). In this paper, we focus on supporting in situ information capturing and report on a realistic laboratory-based study involving experienced paramedics that we used to explore the specific requirements and constraints of supporting in situ information capturing. Specifically, we focused on ways to use audio and visual data capture methods and how they need to be designed to better support paramedics without interfering with their work. We then use the resulting information centric perspective to argue for a roadmap towards smart emergency medical services.

Keywords. Emergency medical service, information systems, emergency communication, emergency documentation, EMS, paramedics

1. Introduction

Emergency medical service (EMS) providers play a crucial role in identifying, treating and transporting patients suffering from time-critical emergency medical conditions. In particular when following the Franco-German model of EMS delivery, patients are treated on site which requires complex clinical judgements [1]. In this context, EMS providers perform urgent actions such as the initial assessment of patients in need and treatment as required (vital signs, intravenous fluids, cardiac monitoring, and stabilization). The decision making leading to interventions requires comprehensive medical knowledge of the physiology and pathology of the human body as well as any pharmacological knowledge of drugs to be administered. Appropriate documentation is required, including documenting the patient's initial condition, the care provided by first responders and EMS providers, the status of the patient during the ambulance transport and responses to any treatment [1,2]. A number of studies [3-5] identified a range of problems that are likely to occur over the course of this complex and time-sensitive process and that may impact on the quality and completeness of the information collected. Common mistakes include insufficiently communicated details of the medical intervention on-site [3-5] or an undocumented initial condition of the patient [6].

¹ Corresponding Author: Kerstin Denecke, Bern University of Applied Sciences, Institute for Medical Informatics, Quellgasse 21, Biel, Switzerland; E-mail: kerstin.denecke@bfh.ch.

Elements such as prehospital hypotension, Glasgow Coma Scale score, and other prehospital vital signs [4] or key elements of the history (witnesses, predisposing factors, and post-event symptoms) and physical examination [5] are not recorded consistently. Some of these problems can be attributed to the fact that EMS personnel may have to remember key information in the above-mentioned stressful setting and only gets to document information once a mission is completed or even towards the end of their shifts [7]. To address these issues, digital protocols become increasingly available in EMS, for example the Medical-Pad provided by Tech2Go (<https://www.tech2go.de>). Some of these systems allow for automatic data capturing from medical devices such as ECG. Speech recognition technology [8] and the use of body worn cameras [9] have been explored, but research in this field is still in its infancy.

In this paper, we report on a realistic laboratory-based study involving experienced paramedics that we used to explore the specific requirements and constraints of supporting in situ information capturing. We present our study design and discuss our findings. We also use the resulting information centric perspective to argue for the need for a roadmap towards smart emergency medical services, and conclusions regarding possible concerns to be addressed.

2. Exploring Options for Real-Time Data Capturing during EMS Delivery

The Medical Informatics *Living Lab* at the Bern University of Applied Sciences in Biel, Switzerland brings together expertise in a range of relevant IT (information systems, process design, human centred design), clinical experiences as well as substantial, hands-on EMS experience. We collaborated with local EMS providers to set up a laboratory based study that helps us to probe the suitability of different in situ data capture methods (audio, visual) and how they could be designed in such a way that they fit paramedics' ways of working without interfering too much with their work practices.

2.1. Experimental Setup

The study was designed around medical emergencies using SimMan, an advanced patient model that can be programmed to simulate medical cases of varying complexity. Two conditions requiring interventions were designed, anaphylactic shock (scenario 1) and a low-back injury (scenario 2). Visible injuries and pain zones were marked on the SimMan by images or text. A skilled research assistant was present on site and adapted the vital signs of SimMan during the study as to simulate the patient's responses to treatments administered. A mission was considered to be completed once the "patient" was stabilized and considered ready for transport by the paramedic team. Each of the teams of paramedics had to handle both cases in the same order with a 10 minutes break in between. We implemented the Wizard-of-Oz method [10], a method used for evaluating technology, interaction methods, and technology acceptance without the need to put the actual technology in place. In this particular study, the Wizard-of-Oz method enabled paramedics to document a situation or a particular aspect thereof (e.g. the visual impression of a wound) by using voice commands to "activate" in situ data capturing. The corresponding audio or visual recording was then administered by the research assistant standing in for the actual technology. Once the teams completed each of the two EMS deliveries, paramedics and researchers met to review and discuss the video recordings of the earlier EMS missions. In these reviews, participants were encouraged

to discuss their use of the "live" documentation feature and also whether they would have preferred different types of recordings (different angle, duration, close-up, wide-angle) or even different media (e.g. a voice recording instead of video recording or video recording where a photograph was taken).

2.2. Data Collection

We recruited as participants four experienced, active, male paramedics between 32 and 41 years old (average 36.5 years) from EMS providers based in the Swiss Canton of Bern. Their work experience ranged between 4 and 12 years (average 8.25 years). Following the Franco-German model of EMS delivery they were to form teams of two. Participants were made aware that the experimental setup was realistic in the sense that be provided standard EMS equipment and that fully engaging in the EMS delivery would likely lead to physical exertion and physical stressors. Participants wear body worn GoPro cameras attached at chest level that documented the EMS delivery roughly from their point of view. As explained earlier, participants could also, at any point in time, "activate" documenting by picture/video/audio recording. This feature did not use the body worn cameras but would allow "free form" data capturing as per participants' specifications. Data collected included number of audio/video/photo recordings requested during the course of the live experiment, the number of audio/video/photo recordings requested while watching the footage, reasons for requesting media as recalled by participants, additional data needed (e.g. data transfer from medical devices such as ECG). Data regarding live recordings was then compared to data recording requests as they emerged during the review.

3. Data analysis and interpretation

In terms of duration it took the two teams between 6.85 and 8.50 minutes to have "patients" suffering from anaphylactic shock and a low-back injury stabilized and ready for transport which is in line with what our own EMS experienced personnel who designed the scenarios expected. The following data requests were made during the course of the experiments (see Figure 1): Audio recordings were requested most ($n=13$), followed by taking a picture ($n=8$). Some video recordings were taken during the mission ($n=3$) to document the overall situation. Participants asked only rarely ($n=1$) for transmitting data via external interfaces (e.g. vital parameters from monitoring systems to the documentation record). When reviewing the footage captured during the live events the paramedics would have preferred to use different data capturing methods: In general, more audio recordings were requested ($m=22$) when reviewing the recording than were requested during EMS delivery. Some video recordings were replaced by other media during the review ($m=2$). While watching the recording they asked more for data collection via external interfaces ($m=8$). The scatterplot (see Figure 2) shows that audio recordings were mainly used for documenting the medical history and administered medications during the mission and afterwards. Pictures were largely requested for documenting wounds and for medical documentation (for example, the documentation dossier of the nursing service). The video recordings were exclusively requested for recording the overall situation.

We looked into reasons why there were these differences between Wizard-of-Oz enabled data capturing during the course of the live event and changes to the way

recordings were made during the retrospective review. One reason is that participants realized that audio recordings would not only help them remember certain aspects but that audio recordings could also be used to transfer data directly to an electronic version of the emergency protocol, via speech-to-text processing, therefore replacing the need for manual data entry. Paramedic participants explained that sometimes, it is difficult to verbalize an emergency situation in a way that all parties can understand the overall situation and surroundings. A video recording could help and was requested in the live event. However, during discussions, participants changed their opinion and preferred taking pictures of the overall situation instead of the initially requested video recording. A reason given for this is that pictures might be more meaningful and, more importantly, provide relevant information at a glance without having to watch a video. During the course of the live event, vital values were documented using audio (paramedics reading the values aloud) or photos (taken of medical devices). During the review paramedics would have preferred to have used audio only if this allowed them to have the data directly entered into an electronic version of the emergency protocol, via speech-to-text processing.

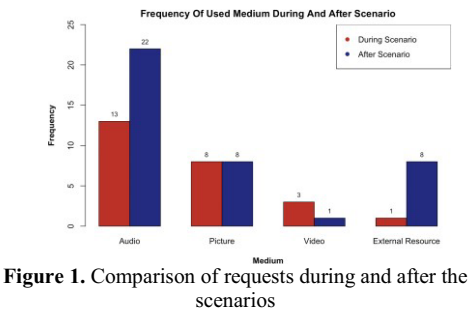


Figure 1. Comparison of requests during and after the scenarios

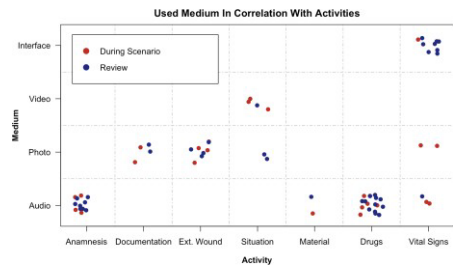


Figure 2. Scatterplot of requests during and after scenario correlated with documentation activities.

4. Discussion and future work

In this paper, we described a study to probe usefulness of different data recording approaches during EMS missions. There is some evidence that incident documentation in EMS could be supported by speech-to-text technology that would allow to enter data directly into the electronic emergency protocol which is needed for documentation purposes and for hand-over to emergency department staff. It is believed that this could help ease the problem of gaps or errors in documentation which are often due to retrospectively collected data. We would expect this to improve both timeliness and correctness of documentation. Ho et al. argue that body worn cameras could help increasing documentation accuracy [9]. While our research supports this observation, we also observed that user acceptance is a critical issue in this context. Participants recognized the benefits of a video recording including increasing their personal security during a mission. However, they were also keenly aware of potential legal ramifications. They know just too well that in emergency situations, decisions regarding treatments may need to be made in a split second and they may include non-standard treatment methods. A potential solution against the fear of after-mission misinterpretations of recordings might be that the recordings are only made available afterwards if paramedics explicitly agree or can add explanations to avoid misunderstandings by not involved

persons. We conclude, there are use cases for video recording in EMS settings, but research is still necessary to find solutions that are accepted by the paramedics and do not hamper the working process.

Given the results of the study, we envision an intelligent rescue environment that supports real-time documentation, direct data capturing from devices, and electronic data transfer to the hospital information system. Further, paramedics have to be equipped with intelligent devices that automatically scan the environment for upcoming risks (like dropping material, changing environment due to weather changes) and provide intelligent decision support as needed, possibly involving services like a telemedicine approach via video guidance [11]. We are currently developing and evaluating a novel system that allows to capture data in a digital EMS protocol using speech recognition technology specifically designed for Swiss EMS and look into its integration with medical devices to enter vital signs into the protocol automatically and to implement data transfer using the CDA-CH RESP standard. In future, we will include methods for ontology mapping (e.g. mapping to a medical ontology [12]) and natural language understanding to make our voice user interface more powerful. Relying upon communication standards and medical ontologies will help in ensuring semantic interoperability. Our research clearly shows the need for acceptance (see also [11]) and that we keep in mind that we don't fall into the trap of designing interfaces that are more suitable for patient, logical, rational humans than for EMS staff who get tired, irritated, or distracted [13].

References

- [1] S. Al-Shaqsi, Models of International Emergency Medical Service (EMS) Systems, *Oman Med J.* **25** (2010), 320–323.
- [2] B. Long, L. Serrano, J. Cabanas, M. Bellolio, Opportunities for Emergency Medical Services (EMS) Care of Syncope, *Prehosp Disaster Med* **31** (2016), 349–52.
- [3] J. Brice, K. Friend, T. Delbridge, Accuracy of EMS-recorded patient demographic data, *Prehosp Emerg Care*, **12** (2008), 187–91. doi: 10.1080/10903120801907687.
- [4] A. Carter, K. Davis, L. Evans, D. Cone, Information loss in emergency medical services handover of trauma patients, *Prehosp Emerg Care* **13** (2009), 280–5.
- [5] B. Long, L. Serrano, J. Cabanas, M. Bellolio, Opportunities for Emergency Medical Services (EMS) Care of Syncope. *Prehosp Disaster Med* **31** (2016), 349–52.
- [6] S. Johnson, Five costly EMS patient documentation mistakes, EMS Legal Update, July 8, 2016, <https://www.ems1.com/ems-products/ePCR-Electronic-Patient-Care-Reporting/articles/105246048-5-costly-EMS-patient-documentation-mistakes/>
- [7] E. Angeli, Three Types of Memory in Emergency Medical Services Communication, *Written Communication* **32** (2015), 3–38.
- [8] S. Blackley, J. Huynh, L. Wang, et al., Speech recognition for clinical documentation from 1990 to 2018: a systematic review, *J Am Med Inform Assoc* **26** (2019), 324–338.
- [9] J. Ho, D. Dawes, E. McKay, et al., Effect of Body-Worn Cameras on EMS Documentation Accuracy: A Pilot Study, *Prehosp Emerg Care* **21** (2017), 263–271.
- [10] P. Green, L. Wei-Haas, The Rapid Development of User Interfaces: Experience with the Wizard of OZ Method, *Proceedings of the Human Factors Society Annual Meeting* **29** (1985), 470–474.
- [11] B. Metelmann, C. Metelmann, Mobile Health Applications in Prehospital Emergency Medicine. In A. Mourtoglou (Ed.), *Mobile Health Applications for Quality Healthcare Delivery*. Hershey, PA: IGI Global, 2019.
- [12] K. Denecke, Semantic structuring of and information extraction from medical documents using the UMLS, *Methods Inf Med* **47** (2008), 425–34.
- [13] C. Lueg, M. Twidale, Designing for Humans, Not Robots (or Vulcans), *Library Trends* **66** (2018), 409–421.

Integrating Pharmacogenetic Decision Support into a Clinical Information System

Kevin TIPPENHAUER^{a,1}, Marwin PHILIPS^a, Carlo R. LARGIADÈ^b, Murat SARIYAR^a, Thomas BÜRKLE^a

^aBern University of Applied Sciences, ^bBern University Hospital (INSEL)

Abstract. Pharmacogenetic testing can prevent adverse drug events but has rarely found its way into clinical routine. One reason is the lack of tools for smooth and automatable integration of pharmacogenetic knowledge into existing processes. Especially, electronic medical records (EMR) represent a suitable environment for such tools. We developed a modular service-oriented prototype of a pharmacogenetic decision support system within an EMR system of the Bern University Hospital. Here, we present the component architecture of our system and discuss issues required for generalizing our results.

Keywords. Pharmacogenomics, pharmacogenetics, decision support, adverse drug events, HL7 FHIR, expert system

1. Introduction

Pharmacogenetics is the study of genetic effects on the metabolic pathways of drugs. About 95% of individuals carry one or more known genetic variants that are important for drug dosing recommendations [1]. Ignoring genetic variation while prescribing a drug can cause adverse drug reactions (ADR) or insufficient drug effects [2–4]. Approximately 20 genes affecting about 80 agents have been identified as actionable in the clinic [5], and the Clinical Pharmacogenetics Implementation Consortium (CPIC) updates continuously related guidelines with dosing instructions for relevant gene-drug combinations [6]. Nevertheless, the actual implementation of pharmacogenetic testing in clinical patient treatment is progressing slowly.

Relling and Evans summarized some general barriers for pharmacogenetic testing in clinical practice: lack of incentives to prevent adverse events; costs and complexity of computational approaches to identify and prioritize genetic variants that influence prescribing decision; disagreement among professional or guideline generating groups; and a lack of awareness of clinicians concerning pharmacogenetic effects. Further, they concluded that the costs associated with pharmacogenetics are shifting from the laboratory testing to the costs for generating evidence-based decisions from the genetic data [5].

Related to existing implementations, Hinderer et al. reviewed 20 pharmacogenetic decision support systems. Sixteen had prototype status and are often built for one

¹ Corresponding Author, Institute for Medical Informatics I4MI, Bern University of Applied Sciences, Quellgasse 21, 2500 Biel, Switzerland; E-mail: kevin.tippenhauer@bfh.ch.

institution. Only one of the productive systems (TreatGx) was integrated with the local EMR [7]. The same lack of EMR-integration is corroborated by the Ubiquitous Pharmacogenomics project (U-PGx), where only one out of the seven involved European countries implemented an active clinical decision support for pharmacogenetics [8]. Having not found any suitable and EMR-related solution, we developed a prototype implementation for assessing gene-drug-combinations in clinical oncology at the Department of Clinical Chemistry of the Bern University Hospital Switzerland (INSEL). Our goal was to derive a component-based approach to support the integration of pharmacogenetic clinical decision support into a broader range of clinical information systems (CIS).

2. Methods

Initially, we identified the main components required for our prototype: inference engine, knowledge base, and data dictionary (see also [9]). Next, we iteratively defined the architecture according to a typical workflow of pharmacogenetic analysis as described by clinicians at the INSEL. We chose the open source tool PharmCAT as the inference engine [10]. It generates pharmacogenetic reports in JSON and HTML from genetic data in the Variant Call Format (VCF) and comes with a set of CPIC guidelines (the clinical knowledge base) in a proprietary JSON format, which we adapted to our needs. To map information about described drugs we used the Swiss medication catalog “hospIndex” [11] which is incorporated in the INSEL EMR system. The link to the EMR user interface was implemented using the HAPI API, an open-source object-oriented HL7 2.x parser for Java [12]. The CDSS user interfaces were created with the proprietary parametrization language of the CIS within the medication module of the INSEL EMR system. In order to verify our prototype, a synthetic data generator for producing VCF data was implemented using the proprietary parametrization language of the INSEL EMR system.

3. Results

The underlying generic system architecture of our solution (Figure 1) considers two different workflows. The first workflow is triggered by the entry of new patient related genetic data and generates a report comprising the relevant pharmacogenetic findings (PGx Report Generator in Figure 1). It uses following components:

- The genetic data repository to store the raw genetic data.
- The knowledgebase with rules to annotate the genetic data.
- The data dictionary to encode the resulting report.
- The genetic annotation engine to generate the pharmacogenetic report.
- The pharmacogenetic report repository to store the report.

When a new genetic record is added to the Genetic Data Repository, the record is pushed to the Genetic Annotation Engine. This engine is supplied with the relevant pharmacogenetic rules from the knowledgebase and with the code mapping (e.g. translating phenotype “Poor Metabolizer” into a LOINC code) from the data dictionary (Knowledge Resources in Figure 1). Finally, the annotated report is added to the shared PGx report repository.

Once a PGx patient report is available (PGx Decision Support System in Figure 1), the second workflow can be triggered by a drug prescription event, which returns the case specific information from the corresponding CPIC guideline. It comprises the following components:

- The CIS, which sends the prescription and receives the decision support result.
- The pharmacogenetic report repository to read the patients report.
- The knowledgebase containing the decision support rules to evaluate the report for the given drugs.
- The data dictionary to encode the results.
- The CDSS-Engine to deduce the correct decision support message.

Upon a new drug prescription, the CIS sends a prescription event with the encoded prescription data to the CDSS-Engine. The CDSS-Engine looks up the encoded drug in the Data Dictionary and loads the corresponding guideline rules from the knowledgebase. If guideline rules are found, the CDSS-Engine tries to find an existing pharmacogenetic report of the patient. If none exists, a warning is sent to the CIS, since a potential pharmacogenetic adverse event cannot be excluded, and a genetic examination might be advisable. If a pharmacogenetic report of the patient exists, the CDSS-Engine applies the guideline rules on the report and returns applicable dosing instructions to the CIS. Finally, the CIS displays the received message to the prescribing physician.

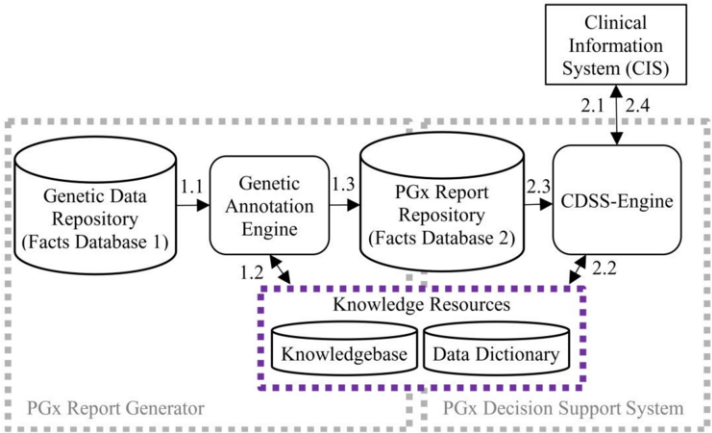


Figure 1. Generic model of a pharmacogenetic decision support system. A Genetic Data Repository acts as the fact database for the Genetic Annotation Engine and provides access (1.1) to the genetic data of the patients. The Genetic Annotation Engine creates a report with the relevant pharmacogenomic findings using the parameters from the shared Knowledge Resources (1.2). The report is stored (1.3) in a PGx Report Repository, which acts as the fact database for a second inference module (CDSS-Engine). Once the rules from the Knowledge Resources have been applied to the patients PGx report, the engine uses the drug prescription from the CIS (2.1) and returns a warning (2.4).

In our implementation, the PGx Report Generator uses a system folder with VCF files for the Genetic Data Repository. A modified version of PharmCAT acts as the Genetic Annotation Engine. The modifications enable PharmCAT to:

- Accept HL7v2 observation request messages as a trigger to read data from the Genetic Data Repository.
- Read the Data Dictionary in form of a configuration file containing the mapping between the genotypes/phenotypes/substances and their encoding.
- Encode genotypes, phenotypes and substances with codes from the Dictionary.

- Turn the pharmacogenetic report into a HL7v2 observation result.

The drug selection event in the CIS oncology drug prescription form triggers the PGx Decision Support System. The CDSS and the PGx Report Repository have been implemented directly within the CIS. The reports from the PGx Report Generator are stored in the database of the CIS acting as the PGx Report Repository for the CDSS-Engine. We implemented the knowledgebase and the data dictionary as internal CIS tables.

For generating synthetic data, we developed a Laboratory Data Simulator (Figure 2). It generates custom VCF data with user defined genetic variants and allows to validate our solution. The following validation workflow was realized in order to compare the issued message with the expected one:

- The validator selects variants to be tested within the laboratory data simulator. The selected variants are added to an existing VCF file template.
- The VCF file is added to the Genetic Data Repository and a HL7v2 observation request message triggers the PGx Report Generator to generate a report.
- The resulting report is returned as HL7v2 observation result to the Pharmacogenetic Report Repository of the CIS and the PGx CDSS is triggered.

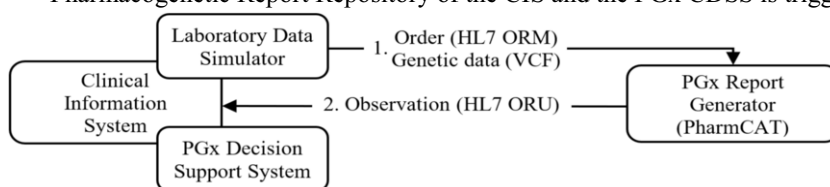


Figure 2. Our prototype integrated with the Laboratory Data Simulator (LDS). In the Clinical Information System (CIS), the user can invoke the LDS and generate a genetic data set for one of the test patients. When finished, the LDS triggers the PGx Report Generator with a HL7v2 ORM message (1). The PGx Report Generator reads the genetic data of the patient and creates a PGx report in the HL7 ORU format, which is then sent to the CIS (2). The user can now either switch to the medication prescription user interface and test the actual workflow integration or test the PGx Decision Support System output directly in the LDS. The system supports pre-test and post-test alerts.

4. Discussion

We presented a component-based architecture for genetic decision support during medication prescription. We could identify six generic components, namely a Genetic Data Repository, a Genetic Annotation Engine, a PGx Report Repository, a CDSS-Engine, a Knowledgebase for PGx rules and a Data Dictionary (see Figure 1).

Due to the following reasons, we consider an architecture with two different workflows (writing genetic reports and providing decision support) as mandatory: First, discussions with physicians revealed that access to the pharmacogenetic report is required to assess the content of the CDSS warnings. Second, the processing of large genetic data records would lead to increased response times without a report in which the drug related phenotypes are available. This makes our approach suitable for such large genetic data contexts.

The modularity of the system enables, a hospital to select components from the system architecture according to its requirements and plugging them together. Our prototype implementation does not completely match the proposed architecture. While the PGx Report Generator is indeed loosely coupled with the CIS using the HL7v2

standard, the PGx CDSS itself depends not only on the CIS in use, but also on its configuration within the hospital. In addition, we have not implemented shared knowledge resources. Instead, the PGx report generator and the PGx decision support system have each their own copy of the knowledge resources, which need to be updated simultaneously to ensure the correctness of the results. Such a solution is vulnerable to inconsistencies. To avoid this risk, the knowledge resources should be shared as depicted in Figure 1.

Using proprietary data formats, as we did to access the knowledge resources reduces the systems adoptability. However, the goal should be to provide a plug and play experience for the whole pharmacogenetic decision support system, as proposed, e.g. by the SMART Health IT movement [13]. Hence, a standardized data exchange model such as HL7 FHIR with the modules “Terminology” and “Clinical Reasoning” should be used for interfacing with the knowledge resources.

Acknowledgement: This work was supported by swissuniversities and Gebert Rüf Stiftung through the program «BREF – First Ventures».

References

- [1] W. Bush, D. Crosslin, A. Owusu-Obeng, J. Wallace, B. Almoguera, M. Basford, et al., Genetic variation among 82 pharmacogenes: The PGRNseq data from the eMERGE network, *Clin Pharmacol Ther* **100** (2016), 160–169. doi:10.1002/cpt.350.
- [2] L. Wang, H.L. McLeod, and R.M. Weinshilboum, Genomics and Drug Response, *N Engl J Med* **364** (2011), 1144–1153. doi:10.1056/NEJMra1010600.
- [3] U.A. Meyer, U.M. Zanger, and M. Schwab, Omics and Drug Response, *Annual Review of Pharmacology and Toxicology* **53** (2013), 475–502. doi:10.1146/annurev-pharmtox-010510-100502.
- [4] U.A. Meyer, Pharmacogenetics – five decades of therapeutic lessons from genetic diversity, *Nat Rev Genet* **5** (2004), 669–676. doi:10.1038/nrg1428.
- [5] M.V. Relling & W.E. Evans, Pharmacogenomics in the clinic, *Nature* **526** (2015), 343–350. doi:10.1038/nature15817.
- [6] CPIC Guidelines, (n.d.). <https://cpicpgx.org/guidelines/>.
- [7] M. Hinderer, M. Boeker, S.A. Wagner, M. Lablans, S. Neue, J.L. Hülsemann et al., Integrating clinical decision support systems for pharmacogenomic testing into clinical routine - a scoping review of designs of user-system interactions in recent system development, *BMC Medical Informatics and Decision Making* **17** (2017), 81. doi:10.1186/s12911-017-0480-y.
- [8] K. Blagec, R. Koopmann, M. Crommentuijn-van Rhenen, I. Holsappel, C.H. van der Wouden, L. Konta et al., Implementing pharmacogenomics decision support across seven European countries: The Ubiquitous Pharmacogenomics (U-PGx) project, *J Am Med Inform Assoc* **25** (2018), 893–898. doi:10.1093/jamia/ocy005.
- [9] T. Bürkle, H-U. Prokosch & J. Dudeck, Steps towards Integration of Knowledge Based Functions into a Hospital Information System, *Studies in Health Technology and Informatics* **34** (1996), 286–290.
- [10] pharmpcat.org, PharmCAT. (n.d.). <http://pharmpcat.org/>.
- [11] HCI Solutions AG - hospINDEX, Daten für Spitäler, (n.d.). <http://www.hcisolutions.ch/de/datenbanken-und-software/index-datenbanken/hospindex.php>.
- [12] HAPI – The Open Source HL7 API for Java, (n.d.). <https://hapi.fhir.github.io/hapi-hl7v2/>.
- [13] SMART Health IT, (n.d.). <https://smarthealthit.org/>.

Localisation, Personalisation and Delivery of Best Practice Guidelines on an Integrated Care and Cure Cloud Architecture: The C3-Cloud Approach to Managing Multimorbidity

George DESPOTOU^{a,1}, Gokce B. LALECI ERTURKMEN^b, Mustafa YUKSEL^b, Bunyamin SARIGUL^b, Pontus LINDMAN^c, Marie-Christine JAULENT^d, Jacques BOUAUD^{d,m}, Lamine TRAORE^d, Sarah N. LIM CHOI KEUNG^a, Esteban DE MANUEL^e, Dolores VERDOY^e, Antonio DE BLAS^f, Nicolas GONZALEZ^f, Mikael LILJAⁿ, Marie SHERMAN^g, Malte VON TOTTEBEN^h, Marie BEACHⁱ, Christopher MARGUERIEⁱ, Liran KARNI^j, Gunnar O. KLEIN^j, Dipak KALRA^k, Rong CHEN^l, and Theodoros N. ARVANITIS^a

^a*Institute of Digital Healthcare, WMG, University of Warwick, Coventry, UK*

^b*SRDC Software Research Development and Consultancy Corp, Ankara, Turkey*

^c*Medixine, Finland*

^d*Inserm, Sorbonne University, University of Paris 13, LIMICS, France*

^e*Kronikgune, Institute for Health Services Research, Spain*

^f*Osakidetza, Spain*

^g*Region Jämtland Härjedalen, Sweden*

^h*empirica Gesellschaft für Kommunikations Technologieforschung mbH, Germany*

ⁱ*South Warwickshire NHS Foundation Trust, UK*

^j*Örebro University School of Business, Informatics, Örebro, Sweden*

^k*European Institute for Innovation through Health Data, Belgium*

^l*Cambio Health Systems, Sweden*

^m*AP-HP, Delegation for Clinical Research and Innovation, Paris*

ⁿ*Department of Public Health and Clinical Medicine, Unit of Research, Education, and Development Östersund Hospital, Umeå University, Umeå, Sweden*

Abstract. *Background:* C3-Cloud is an integrated care ICT infrastructure offering seamless patient-centered approach to managing multimorbidity, deployed in three European pilot sites. *Challenge:* The digital delivery of best practice guidelines unified for multimorbidity, customized to local practice, offering the capability to improve patient personalization and benefit. *Method:* C3-Cloud has adopted a co-production approach to developing unified multimorbidity guidelines, by collating and reconciling best practice guidelines for each condition. Clinical and technical teams at pilot sites and the C3-Cloud consortium worked in tandem to create the specification and technical implementation. *Results:* C3-Cloud offers CDSS for diabetes, renal failure, depression and congenital heart failure, with over 300 rules

¹ Corresponding Author, George Despotou, Institute of Digital Healthcare, WMG, University of Warwick, Coventry CV4 7AL, United Kingdom; E-mail: G.Despotou@warwick.ac.uk

and checks that deliver four best practice guidelines in parallel, customized for each pilot site. *Conclusions:* The process provided a traceable, maintainable and audited digitally delivered collated and reconciled guidelines.

Keywords. multimorbidity, integrated care, clinical interpretable guidelines, co-production, clinical decision support

1. Introduction

There is an increasing need to organize care around the patient and not on conditions [1], taking into account his or her multiple physical and psychosocial conditions [2]. An integrated, patient-centered care and cure delivery architecture needs to be developed considering the realities of multi-morbidity and poly-pharmacy. This needs to take into account the medical, technological, organizational and socio-economic challenges of creating a collaboration environment for all of the stakeholders involved in the holistic continuum of care. Clinical guidelines re-factor the best available published evidence on clinical effectiveness, into decision trees and care pathways. They are increasingly multi-professional, and work across care providers. However, they usually focus on treating a single disease. Clinical guidelines may clash (e.g., due to incompatible treatment and monitoring regimes). Following more than one clinical guideline, can result in inefficiencies for the patient and for the health system, due to duplicating and inconveniently scheduled investigations and clinic visits, and more importantly, treatments that may exacerbate another condition [3].

C3-Cloud is ICT infrastructure enabling a collaborative care and cure cloud for continuous coordination of patient-centered care activities; by a Multidisciplinary care Team (MDT), and patients/informal caregivers. It implements integrated care pathways for chronic heart failure, renal failure, depression and diabetes, based on best practice guidelines [4]. The C3-Cloud approach to integrated care is demonstrated in varying clinical, technological and organizational settings by piloting in three European regions (South Warwickshire, Basque Country, and Region Jämtland Härjedalen) with considerably different health and social care systems, and ICT landscapes. The paper presents the co-production, and implementation approach adopted in C3-Cloud, which delivers an infrastructure able to offer multiple personalize-able best practice guidelines for multimorbidity. C3-Cloud also incorporates localization, necessary to be deployed in the heterogeneous pilot site environments.

2. The Collaborative Care and Cure (C3-Cloud) Solution to Multimorbidity

C3-Cloud achieves its objectives using the concept of the personalized care plan. The care plan consolidates and reconciles data, goals, treatment, and medications, personalized for each patient, addressing their collective multimorbidity needs, and encapsulating clinical practice guidelines. The care plan can be thought of as an integrated care visualization and management tool. Figure 1 presents the high-level architecture of C3-Cloud and its constituent components, which offer the care plan. The Coordinated Care & Cure Delivery Platform (C3DP) is an innovative online means for multidisciplinary care team members (MDT) to collaboratively manage (execute, monitor, update) the integrated personalized care plans for patients with multi-morbid

conditions. Clinical Decision Support Services (CDSS) support personalized care plan development and execution by clinical guideline reconciliation, risk stratification, poly-pharmacy management, and goal setting and monitoring. A Patient Empowerment Platform (PEP) ensures active participation of patients and their informal caregivers to the management of their multi-morbid chronic conditions, thus alleviating the non-adherence problem. Interoperability middleware addresses technical (TIS), semantic (SIS), and privacy/security (SPS) interoperability challenges to seamlessly integrate with the existing health care, social care, and home/community care information systems. This will allow interfacing the information exchanges, but also the clinical terminology systems between the local systems and C3-Cloud.

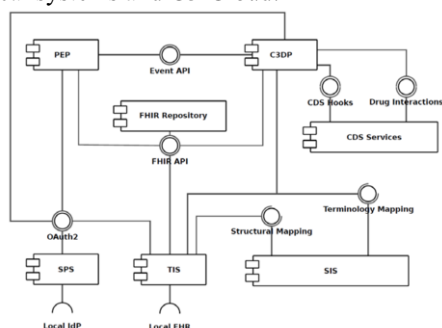


Figure 1: Overview of the C3-Cloud Architectural Components

Finally, all data are stored in a database as HL7 Fast Healthcare Interoperability Resources (FHIR) (<https://www.hl7.org/fhir/>), annotated with the codes of the terminology systems, C3-Cloud uses. FHIR, along with other standards used internally in C3-Cloud, result in an easily maintainable and upgradeable architecture.

3. From Medical Knowledge to Localization and Personalization

One of the main objectives of C3-Cloud is to deliver integrated care, based on best practice guidelines. In order to achieve this, C3-Cloud had to overcome a number of challenges: a) unify and reconcile guidelines to offer the patient a single multimorbidity management plan rather than a digitization of multiple separate guidelines; b) deliver an infrastructure suitable to customize the vision of an enhanced service for all pilot sites; and c) increase patient benefit by offering further personalization options based on what the clinical team of each pilot site identified as important. Figure 2 illustrates a process that enabled co-production of the consolidated and reconciled multimorbidity guidelines delivered by C3-Cloud. Local practice for each pilot site was identified by the clinical teams of each pilot site. Then, pilot sites recorded their vision of an enhanced service, and performed a gap analysis with the current service to identify core functional requirements for C3-Cloud. This includes personalization rules that need to be applied in multimorbidity, which are then translated into service processes. For example, online messaging to adjust goals of a patient, and online patient reported outcomes triggering automated clinical recommendations.

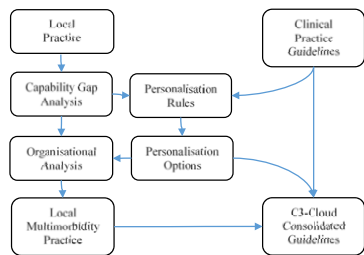


Figure 2: C3-Cloud Consolidated Multimorbidity Guidelines Process

An organizational analysis was then performed to identify how the newly defined capability would be integrated to each pilot site, including allocation of roles, and specification of local procedures. Figure 3 illustrates a comparison of how the C3-Cloud enabled service would fare against existing practice, as evaluated by clinical teams in each pilot site. The evaluation dimensions were identified by the C3-Cloud consortium based on current literature on integrated care. This process was informed by the UK National Institute of Care Excellence (NICE) evidence-based guidelines, which provided a skeleton gold standard that all pilot sites used as reference. NICE guidelines balance best available evidence with cost benefit, ultimately resulting in a gold standard for care. Nevertheless, even the NICE guidelines, offer limited supported to multimorbidity.

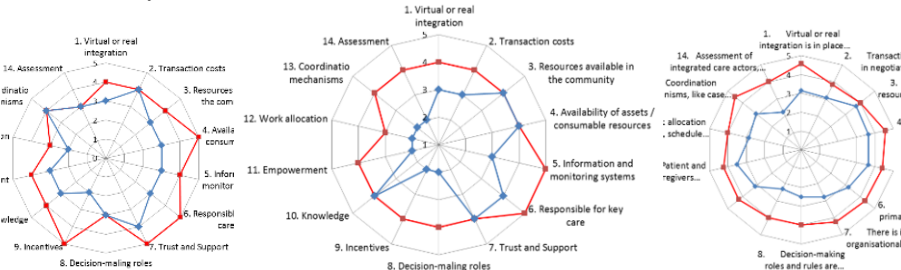


Figure 3: Before and after pilot site internal clinical evaluation of the service

Multimorbidity options, rules and actions, were developed by the C3-Cloud Clinical Reference Group (CRG), consisting of clinicians from each pilot site. The CRG also had overall oversight of the resultant clinical practice. Throughout this process, functional and technical requirements for C3-Cloud were also identified by using a Model-Driven Engineering approach, which resulted in the specification of the software architecture, as well as the system specification document.

4. Implementation of Guidelines in C3-Cloud

The specification of the C3-Cloud guidelines was followed by their implementation, and deployment as part of the system. The guidelines development process resulted in a number of technical requirements for C3-Cloud. They provided requirements such as: data that needed to be stored in the FHIR database; the user interface that users (MDTs and patients) would experience, capturing all aspects of multimorbidity management identified in the guidelines; and automated rules implemented by the clinical decision

support modules. Figure 4 presents the steps followed for implementation of the CDSS logic, and integration with the local sites.

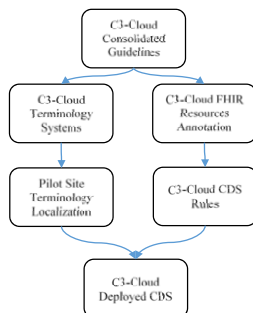


Figure 4: Guidelines Implementation and Deployment in C3-Cloud

In order to meet the interoperability requirements, each step of the C3-Cloud guidelines was annotated with terminologies by the CRG. This allows for unambiguous retrieval of the relevant information from the FHIR repository, as well as an automated interface, which mapped the C3-Cloud terms to the terms used in each pilot site. Furthermore, the guidelines were annotated with the FHIR resources with which they needed to interact. For example, when patients enter data, the system will create an [Observation] FHIR resource, and when the CDSS identifies suitable conditions for certain medications (e.g. beta blockers for high blood pressure), the system will propose to a physician new medication, by creating a [MedicationRequest] resource that they will need to review before approving. This resulted in an annotated logic of the guidelines, that was then implemented using the GDL2 openEMR standard [5]. The GDL2 implementation is deployed as an executable CDSS module, in each pilot site.

5. Conclusions

C3-Cloud had deployed collated and reconciled guidelines on multimorbidity in three European pilot sites. Achieving this required a co-production process involving the pilot site clinical teams, the C3-Cloud clinical reference group, the pilot site technical teams, and the C3-Cloud technical team. The process resulted in over 60 clinical scenarios, 300 CDS rules and over 380 design requirements.

References

- [1] The King's Fund, 'Providing integrated care for older people with complex needs: Lessons from seven international case studies'. <http://www.kingsfund.org.uk/publications/providing-integrated-care-older-people-complex-needs>.
- [2] U. Kadam, Redesigning the general practice consultation to improve care for patients with multimorbidity, *BMJ* **345** (2012), e6202, doi: 10.1136/bmj.e6202.
- [3] B. Guthrie, K. Payne, P. Alderson, M.E.T. McMurdo, S.W. Mercer Adapting clinical guidelines to take account of multimorbidity, *BMJ* **345** (2012), e6341, doi:10.1136/bmj.e6341.
- [4] Erturkmen GBL, Yuksel M, Sarigul B, Arvanitis TN, Lindman P, Chen R, *et al.*, A Collaborative Platform for Management of Chronic Diseases via Guideline-Driven Individualized Care Plans, *Computational and Structural Biotechnology Journal* **17** (2019), 869-885.
- [5] Guideline Definition Language v2, Technical specifications (2018). <https://github.com/gdl-lang/gdl2>

Making a Subjective Notion Computer-Interpretable: The Case of the Tumour-Volume to Breast-Volume Ratio for the Surgical Decision of Breast Cancer

Claudia BLANCAFORT^{a,1}, Jacques BOUAUD^{b,a}, Coralie PREBET^c, Brigitte SEROUSSI^{a,c}

^a Sorbonne Université, Université Paris 13, Sorbonne Paris Cité, INSERM, UMR S_1142, LIMICS, Paris, France

^b AP-HP, DRCI, Paris, France

^c AP-HP, Hôpital Tenon Paris, France

Abstract. Clinical practice guidelines (CPGs) often include ambiguous criteria making their translation as computer-interpretable guidelines a difficult task. In breast cancer management, whether to perform a breast conservative surgery (BCS) or not is one example. Most international CPGs recommend to perform a BCS when the tumour volume / breast volume ratio allows for good cosmetic results, which cannot be directly translated into a computable format. We propose to compute an estimate of the ratio using the maximum size of the tumour to compute the tumour volume and the bra size to compute the breast volume. In addition, we take into account the location of the tumour according to quadrants and unions of quadrants. The model has been tested on a retrospective sample of 34 clinical decisions of a breast cancer unit in a Parisian university hospital (France). Concordance was found in 91.2% of the cases, with good sensibility and specificity. This finding could set a new pathway to advance on the development of actionable decision criteria to be used in a future clinical decision support system for breast cancer management.

Keywords. Clinical decision support systems, Breast cancer surgery, Tumour burden, Clinical practice guidelines, Computer interpretable guidelines

1. Introduction

Although breast cancer happens to be one of the most advanced research fields in oncology, each step achieved implies an increase on management complexity. Professional organisations and expert groups regularly publish clinical practice guidelines (CPGs) in an effort to bring state-of-the-art knowledge to the medical community. Despite literature clearly reports that CPG compliance leads to a significant improvement in overall survival and disease-free survival [1], adherence to CPG recommendations is lower than expected [2]. The development of

¹ Corresponding Author : E-mail : cblancafortgc@gmail.com

multidisciplinary breast cancer units (BCUs) has proven to improve guideline compliance [3,4], but daily practice is hampered by the increasing number of complex patients and data available.

Clinical decision support systems (CDSSs) have been suggested as possible tools to promote the implementation of guideline recommendations. In fact, embedding CPGs in those systems has shown some improvement on overall guideline compliance [5]. In the breast cancer domain, several CDSSs have been developed such as MATE [6], OncoCure [7], OncoDoc [8], or DESIREE [9], but their development remains a challenge. Most CPGs are developed as narrative documents which makes them difficult to be translated into a computer-interpretable format. Clear examples, among others, is the presence of subjective notions, very difficult to be formally modelled, and the reference to decisional criteria which are not routinely collected.

In breast cancer management, the indication of conservative surgery (BCS) against radical surgery is one of the concepts harder to translate. Normally, the goal of surgery is to ensure both local control and breast preservation with an optimal cosmetic outcome. The current standard of care is BCS followed by adjuvant treatment which is equivalent, in terms of overall survival, to a mastectomy. CPGs propose several criteria contraindicating BCS, some of them being objective such as inflammatory breast cancer, contraindications to radiotherapy, or the patient's choice. Apart from these well-defined situations, CPGs recommend to perform BCS if "optimal surgery is feasible, maintaining good cosmetic results". Then, for most cases the decision whether to perform a BCS is subjective, making it difficult to formalize this criterion in order to use it in a CDSS. The Erasmus MC Cancer Institute research group published a series of works dealing extensively with this issue. After determining the three independent predictors of superior cosmetic result after BCS, tumour volume (TV), breast volume (BV), and tumour location, they published a prediction model based on the TV/BV ratio [10,11]. Unfortunately, the direct application of this model remains complicated since neither TV nor TB is part of the routine workup for breast cancer.

In this paper, we propose an adaptation of this prediction model to make it more easily usable, especially for automated CDSSs. Considering that the tumour location is available in patient records, the aim is to compute the TV/BV ratio from an assessment of both tumour size and patient bra size and to translate the notion automating the recommendation of either BCS or mastectomy. In order to evaluate if this approach is concordant with current clinical practice, the model has been tested on a retrospective sample of cases.

2. Material and methods

2.1. Tumour volume

Measurement of a breast tumour volume normally implies complex techniques such as MRI or 3D ultrasounds. These costly techniques are not systematically used in an ordinary breast cancer extension workup. We considered to estimate the tumour volume assuming the tumour to be a sphere. Since the maximum tumour size is essential information in any breast cancer record, we used it as the diameter of the sphere, thus overestimating the tumour volume.

2.2. Breast volume

Even though breast volume can be assessed by MRI and some complex formulas from mammography measurements, still it is not a standard practice yet. Thus, we proposed to extrapolate the breast volume from the bra size of the patient. Bra size is not systematically described in the medical records, but is it is quite commonly used to describe breast size in BCUs. The relation between bra size and breast volume in cm3 was obtained from a technical pattern and sewing association [12], which is used as a manufacturer design standard in France/Spain/Belgium and is reported in Table 1.

Table 1. Approximate bra size-volume equivalents

Volume (cm³)	Contour and cup					
200-250	80A					
250-300	80B	85A				
350-400	80C	85B	90A			
450-500	80D	85C	90B	95A		
550-600	80E	85D	90C	95B	100A	
650-700		85E	90D	95C	100B	105A
750-800			90E	95D	100C	105B
850-900				95E	100D	105C
950-1000					100E	105D
1050-1100						105E

2.3. Cutting points and tumour location

According to the Lagendijk et al. model [11], cutting points below which BCS is expected to achieve a satisfactory cosmetic result are when the TV/BV ratio are below 21.6 in case the tumour is located in the upper lateral quadrant, 4.1 in the lower lateral quadrant, 15.1 in the upper medial quadrant, 3.2 the in lower medial quadrant, and 15.7 in the central position.

Again, a limitation was encountered, since in a substantial number of cases, the location of the tumour is described in the union of quadrants. Thus, we extended the model and assumed that the cutting points in those junctions of quadrants could be estimated by the arithmetic mean value of the cutting points of adjacent quadrants, resulting as it follows: 18.35 for the upper quadrants junction, 3.65 for the lower quadrants junction, 12.85 for the lateral quadrants junction, and 9.15 for the medial quadrants junction.

2.4. Assessment of the method

We performed a retrospective analysis to assess the method on a sample of clinical cases. The study population consisted of female breast cancer patients submitted to BCU meetings between August 2019 and October 2019 at Hôpital Tenon, a university hospital part of Assistance Publique–Hôpitaux de Paris (AP-HP), in France. All patients had given their informed consent to the re-use of their pseudonymised data for research purpose at hospital admission. Cases where collected in October 2019 when the following inclusion criteria were met: (1) Pre-malign or malign breast tumour confirmed by histological evaluation, (2) Radiologic extension study available, (3) Bra

size specified, (4) Decision regarding surgical management made. Patients were excluded in case of bilateral disease, separated bifocal tumours, prior neoadjuvant therapy, or surgical treatment affected by patient's preferences.

3. Results

Among nine BCU meetings, 174 clinical cases were analysed. From 34 to 93 year-old, the average age was 55.9. Both unifocal and close bifocal tumours were taken into account with an average size of the whole tumour mass of 12.8 mm. Eighty-seven medical records specified the bra size (missing data in 50% of the cases). When the bra size was available, the mean breast volume was 556 cm³, which corresponds to a majority of 90C and 95B. After careful review, 34 cases met the inclusion criteria. BCU physicians chose mastectomy (radical surgery) in 26.5% of the cases, whereas the 73.5% of cases left where oriented towards a lumpectomy (BCS).

For each case, TV/BV ratio was calculated and, according to the tumour location, was classified to determine the surgical procedure recommended by our extended version of the Lagendijk et al. model. Then, each recommendation was confronted with the actual decision made by BCU physicians. Concordance reached 91.2%. Model sensibility and specificity were 88% and 100%, respectively; positive and negative predictive values were 100% and 75% respectively.

4. Discussion

Breast volume is not routinely computed pre-operatively because as yet there is no accepted technique to assess its value. Bra size represents an accessible information since most women know it or it can be assessed by simple anatomic measurements. However, there is not much published in medical literature about the relation between breast volume and bra size. On the other side, this relation has been studied in other domains such as sewing and patronage or even plastic surgery prothesis manufacturers. We have found interesting to exploit those sources of knowledge in profit of science.

According to our results, bra-size is recorded by clinicians in 50% of the cases discussed during BCU meetings. Whether this size is obtained by clinical history or by anatomic measurements could not be assessed. Of course, bra size does not characterize breast shape or ptosis, and it has been suggested that more than half of woman population could be wearing a wrong size of bra [13]. Nevertheless, our results showed that it can be feasible to estimate the breast volume to be used in the computation of the TV/BV ratio. Our recommendation is to get the bra size during physical examination, measuring chest and bust. Regarding the tumour volume, assuming the tumour shape as a sphere has no evidence-based studies to rely on, but this approximation has been done by other research groups [14]. In any case, the documentation of the maximal tumour size along with the tumour localisation is mandatory for staging breast cancer.

Given the good results for sensibility and specificity, our extended version of Lagendijk et al. predictive model appears to be secure and could be used in CDSSs as an orientation tool to decide between BCS or radical treatment. The 8.8% non-concordant decisions were cases for which the model suggested mastectomy whereas BCS was decided. These decisions concerned low malignancy tumours, which could explain the choice of BCU physicians for a less aggressive treatment. Accuracy of the

volume measurements (TV and BV) were not addressed in this study. The goal was not to develop a new clinical predictive model, which would require a much more developed methodological strategy and a higher number of cases. Hopefully this issue is already being addressed by important and experienced research groups.

The actual lack of more explicit definitions regarding the surgical choice for breast cancer patients makes it mandatory to invent creative solutions. We believe this finding could alleviate the frustration in creating a more precise computer-based CDSS to assist with the management of breast cancer from evidence-based guidelines.

Acknowledgement, The authors acknowledge the Alfonso Martín Escudero Foundation (Madrid, Spain) for its financial support of the first author's grant for research.

References

- [1] Wöckel A, Kurzeder C, Geyer V, Novasphenny I, Wolters R, Wischnowsky M, et al. Effects of guideline adherence in primary breast cancer--a 5-year multi-center cohort study of 3976 patients. *Breast* **19**/2 (2010), 120–127.
- [2] Kiderlen M, Ponti A, Tomatis M, Boelens PG, Bastiaannet E, Wilson R, et al. Variations in compliance to quality indicators by age for 41,871 breast cancer patients across Europe: a European Society of Breast Cancer Specialists database analysis. *Eur J Cancer* **51**/10 (2015), 1221–1230.
- [3] Kesson EM, Allardice GM, George WD, Burns HJ, Morrison DS. Effects of multidisciplinary team working on breast cancer survival: retrospective, comparative, interventional cohort study of 13 722 women. *BMJ* **344** (2012), e2718.
- [4] van Hoeve J, de Munck L, Otter R, de Vries J, Siesling S. Quality improvement by implementing an integrated oncological care pathway for breast cancer patients. *Breast* **23**/4 (2014), 364–370.
- [5] Jaspers MW, Smeulders M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc* **18**/3 (2011), 327–334.
- [6] Patkar V, Acosta D, Davidson T, Jones A, Fox J, Keshtgar M. Using computerised decision support to improve compliance of cancer multidisciplinary meetings with evidence-based guidance. *BMJ Open* **2**/3 (2012), e000439.
- [7] Eccher C, Seyfang A, Ferro A. Implementation and evaluation of an Asbru-based decision support system for adjuvant treatment in breast cancer. *Comput Methods Programs Biomed* **117**/2 (2014), 308–321.
- [8] Seroussi B, Bouaud J, Antoine EC. OncoDoc, a successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer. *Artif Intell Med* **22**/1 (2001), 43–64.
- [9] Seroussi B, Lamy J-B, Muro N, Larburu N, Sekar BD, Guézennec G, et al. Implementing Guideline-Based, Experience-Based, and Case-Based Approaches to Enrich Decision Support for the Management of Breast Cancer Patients in the DESIREE Project. *Stud Health Technol Inform* **255** (2018), 190–194.
- [10] Vos E, Koppert L, van Lankeren W, Verhoef C, Koerkamp BG, Hunink M. A preliminary prediction model for potentially guiding patient choices between breast conserving surgery and mastectomy in early breast cancer patients; a Dutch experience. *Quality of Life Research* **27** (2018), 545–553.
- [11] Lagendijk M, Vos EL, Koning AHJ, Hunink MGM, Pignol JP, Corten EML et al. Tumor-volume to breast-volume ratio for improving cosmetic results in breast cancer patients. *BMC Cancer* **17**/1 (2017), 336.
- [12] Medidas y volumen de las copas de seno [internet]. 2014. CC1C. Barcelona. Available at: <http://www.comocubruncuerpo.org/medidas-y-volumenes-de-las-copas-de-seno/>
- [13] White J, Scurr J. Evaluation of professional bra fitting criteria for bra selection and fitting in the UK. *Ergonomics* **55**/6 (2012), 704–711.
- [14] Lagendijk M, Vos EL, Ramlakhan KP, Verhoef C, Koning AHJ, van Lankeren W, et al. Breast and Tumour Volume Measurements in Breast Cancer Patients Using 3-D Automated Breast Volume Scanner Images. *World J Surg* **42**/7 (2018), 2087–2093.

More Information Continuity Through Health Networks? Barriers to Implementation and Use of Health IT

Özlem TECKERT^{a,1}

^a Faculty of Management, Culture and Technology, Osnabrück University AS, Germany

Abstract. Wound networks, as a voluntary structure formed by interdisciplinary health professionals, have the capacity to improve the continuity of care. Health IT systems (HIT) might be able to support these networks ensuring the information continuity in this process but electronic networking usually fails. The aim of this qualitative study is to examine the barriers why the networks do not take this step and possible approaches on how to overcome them. Based on interviews with 11 managers of wound networks 13 barriers could be detected which can be assigned to the main categories personal, financial, technical and legal barriers. It becomes clear that networks as voluntary bottom-up movements cannot increase the information continuity and hence also not the continuity of care on their own. This requires an overall concept. However, according to health experts, this must be drawn up, financed and above all coordinated by policy makers. In addition, these experts postulate structural guidelines for a cooperation in networks as well as nationally binding processes and standards in wound care. These guidelines and processes should then be mapped in a mandatory central HIT.

Keywords. Continuity of care, health information technology, integrated care, qualitative research

1. Introduction

A vertical or horizontal networking of health professionals (HP) along a care pathway can build an integrative structure to improve the continuity of care (CoC) and the quality of care (QoC) [1-3]. The need for integration is especially given for the increasing number of elderly multi-morbid patients with chronic diseases [4]. The chronic wound is also one of these diseases. In Germany alone, studies demonstrate that more than 890,000 predominantly elderly people suffer from chronic wounds, almost 85% of them suffering multi-morbidity [5]. To better coordinate wound care for the benefit of the patient and to increase the CoC, many HP voluntarily form so-called interprofessional wound networks (IPWN) [2]. According to our research over 90 such IPWNs exist in Germany. However, the number fluctuates due to persistent new foundations and dissolutions.

To ensure the information continuity as part of the CoC [6], the implementation of health IT systems (HIT) in a network is seen as a key enabling factor, which also has directly or indirectly an effect on QoC [1,4,7]. Networks of HP with existing social

¹ Corresponding Author, Osnabrück University of Applied Sciences, Kaiserstr. 10c, 49809 Lingen, Germany; E-mail: o.teckert@hs-osnabrueck.de.

relationships represent a solid basis for the implementation and use of a shared HIT [8]. A first analysis of the wound networks reveals that an IPWN alone is not sufficient for establishing an electronic network for the benefit of the wound patients. To this end, freely accessible online data material on 37 of these IPWNs was compiled and searched for key terms such as electronic, digital or IT using MAXQDA. Only nine networks had information on the topic of electronic networking, which does not mean that a shared HIT is actually employed.

For this reason, this study aims to examine the barriers why these IPWN do not take the next step by trying to increase the information continuity through a shared HIT and possible approaches how to overcome these barriers.

2. Methods

Numerous studies deal with barriers to IT use in the health sector. However, these are usually limited to individual medical specializations or institutions. Furthermore, these studies are suitable as a starting point but do not allow for differentiated insights into IPWN. For this reason, we chose a qualitative approach applying guideline-based expert interviews, prompting network managers to report on their experiences with HIT in IPWN. As these interviews were carried out in the context of an overarching project, we also asked questions on other network-specific topics that are not part of this study.

Selected main contacts (e.g. named as managers, coordinators, speakers, chairmen) of IPWN were written to, contacted by telephone or addressed at wound congresses. All together 11 experts (two physicians, three nurses/wound manager, two medical supplier/pharmacist, three managing directors of a management company with a medical background and one other HP) could thus be recruited. These experts are managing 13 wound networks that are under closer examination. The interviews were conducted face-to-face between April and September 2018 and lasted from 74 to 130 minutes each (Ø 104 minutes).

Nine of the networks are a non-profit association. Four networks do not have their own legal form, of which two are publicly funded project networks. In addition, one network is organized as an ICW² working group. All networks are multiprofessional. With the exception of one network, the connectivity is intersectoral. Only two of the networks label themselves as supraregionally active networks.

All interviews were recorded and transcribed. The evaluation was framed by a deductive-inductive content analysis. For this purpose, we created a deductive category system. This system was inductively expanded throughout the course of the analysis [9].

3. Results

According to the experts, only one network never raised the issue of a HIT application. In the other 12 IPWNs this was at least discussed, and in five networks an electronic patient record (EPR) was finally introduced. In two IPWNs the HIT application was discontinued after the test phase. In another IPWN HIT was not used at all, and in the other two IPWN

² The Initiative Chronische Wunden e.V. [Initiative Chronic Wounds] was launched in 1995 by physicians, nurses and other activists in order to improve the prevention and treatment of people with chronic wounds. URL: <https://www.icwunden.de/da/about-us.html>

the EPR is currently used by at least one network member, even while use falls short of expectations. The experts agree that HIT can support networking in various ways. A number of network activities that go beyond general personal information exchange or joint training activities are only made possible by this. However, what are the reasons for either not implementing a HIT or for a limited use? In the course of the content analysis, we initially defined four main topics inductively as possible categories of barriers: A) personal, B) financial, C) technical and D) legal [10]. The subcodes, as displayed in Table 1, were defined inductively.

Table 1. Barriers of a HIT application: list of defined codes and subcodes

Code	Subcode
A Personal	Heterogeneity of the actors (requirements, demands, goals)
	Doubts about the advantages
	Holding on to the tried and tested
	Fear of control of other network members/restriction of autonomy to act
	Lack of trust in data of other network members
B Financial	Lack of technical affinity and competence
	Investments in hardware and software
	Financial losses due to additional time expenditure
C Technical	Financial losses due to the discontinuation of medical examinations
	Missing interfaces / interoperability
	Missing network coverage
	Missing hardware equipment
D Legal	Lack of technical support

Category A: Personal barriers are mentioned by all experts without exception. The *heterogeneity of the actors* requires different needs, demands and objectives in relationship to HIT. In addition, the HP have *doubts about the advantages* of an EPR. “The system itself, the communication platform itself, does neither unite the network nor does it result in higher quality. If I have not defined quality beforehand, it will not be complied with. It can only be a means to an end and will be accepted wherever I really see it. I have been logging it with a lot of effort so far, and now I press a button and it is easier.” (Managing director 2) Many HP prefer to *hold on to the tried and tested* because a new HIT always means a change in individual working methods and routines. Older network members, in particular, are generally skeptical about new digital applications. A *feared restriction of the autonomy to act* and the *fear of control by other members* “[can] be so negative that people always have the feeling that we are always looking into their cards. And that everyone would rather do their own thing” (Nurse 1). In addition, a *lack of trust in the data of other HP* may prevent the introduction of a HIT. A *lack of technical affinity and competence* also means that for many nurses and physicians the topic is “a book with seven seals” (Managing Director 1). Hence, it receives little attention in voluntary wound networks.

Category B: Financial barriers are among the most frequently mentioned barriers in the literature [10]. The experts allude *investments in necessary hardware and software*. Moreover, a new HIT can also result in running costs, for instance, for support. A possible interface to existing IT systems in order to establish interoperability and avoid double documentation is also described as very costly. HP fear *financial losses due to additional time*, if the new system must be used in parallel with existing systems. Fears of *financial losses due to the elimination of medical examinations* due to electronic exchange are also mentioned. Eventually, double examinations represent a source of income for HP.

Category C: Technical barriers result from a *lack of interfaces or interoperability*. “To dare the next step: I hook up there, just pick up something and can even write something in there. That does not work! The office-based physician asks: how can that enter my practice management software system? Not at all! The hospital doctor in the ambulance asks: how do I recognize it? Neither, because these are all structures that do not belong together, and do not belong together intentionally.” (Managing director 2) Inadequate reception quality - for example in areas with a *low network coverage* - and poor or *missing hardware equipment* on the part of the members make IT networking impossible. A *lack of technical support*, e.g. by a software company, can also delineate a potential usage barrier.

Category D: Legal issues. *Data protection* is a sensitive issue, especially in the healthcare sector. Doubts as to whether highly sensitive patient-related data are secure with other users and are not misused are a potential barrier.

4. Discussion

The exploratory research demonstrates that interconnectedness of HP alone does not pave the way for the implementation and use of health IT. Similarly, none of the networks considered can be highlighted as the best case.

According to the experts, an attempt to overcome the barriers described in the categories A-D separately would be insufficient. This is due to the interconnectedness of the barriers and the difficulty to gain direct impact on them [10]. This can be illustrated by several examples: some of the experts complain about the lacking financial support for a HIT, which impairs the funding of such a project. However, the costs alone cannot determine the lack of software introduction and application. For example, the HIT in one of the IPWN was financed by public project funds, and the use was free of charge for the actors involved. Nevertheless, it was hardly used or completely rejected. Four other experts report a similar situation, where the software and some of the hardware were offered to network members free of charge or for only a small fee. However, in the end the expected success failed to emerge. In the case of all these networks, the HIT was even developed internally or selected together with the members for the sake of diversity of the actors. Despite initial euphoria, training courses to harmonize wound care and increase technical competence did not result in the desired use.

Overcoming the barriers is a complex process and requires a holistic concept and support from various actors. Voluntarily formed networks themselves cannot guarantee information continuity. “After all, a network is also a bottom-up movement, if you like. It is professionally triggered, it is largely care-triggered but does not solve the system problem in this way”. (Managing director 2)

The establishment of a network and the topic of HIT depend too much on the commitment of individual HPs. Once the networking and implementation of HIT succeed, then only isolated solutions are created. The experts see a particular duty on politicians, who should not only provide the financial means for individual pilot projects. They must also accompany and coordinate networking at a regional level in terms of its content. According to the experts, this includes creating discussion platforms and structural guidelines for cooperation and communication in networks and, above all, defining nationally binding processes and standards in wound care that are mapped in a mandatory and central HIT. Instead of an individual obligation to provide evidence, it might also help

to enable an integrative consideration of services. This study is limited due to the small sample size. In addition, the findings are confined to wound networks and cannot be readily generalized to other health networks. Evaluating the impact of the structure and the quality of the networks on HIT usage is subject for future work [11].

5. Conclusion

The study demonstrates that networks of HP cannot guarantee information continuity. The experiences of 11 managers reveal that, despite the voluntary structures created, there are numerous barriers to a HIT implementation and use. The overcoming of these barriers should not be left to the commitment and voluntariness of individual HPs but should be actively controlled by policymakers. In this way, QoC can be improved by HIT used across institutions [3].

Acknowledgements

This study was funded by the Federal State of Lower Saxony Germany (grant: ZN 3062).

References

- [1] A. Wierdsma, C. Mulder, S. de Vries, S. Sytema. Reconstructing continuity of care in mental health services: a multilevel conceptual framework, *J Health Serv Res Pol.* **14** (2009), 52–7.
- [2] R. Sheaff, J. Schofield, Inter-Organizational networks in health care. In: E. Ferlie, K. Montgomery, A. Pedersen, (eds.), *The Oxford handbook of health care management*, OUP, Oxford, 2016.
- [3] A. Pinsonneault, S. Addas, C. Qian, V. Dakshinamoorthy, R. Tamblin, Integrated health information technology and the quality of patient care: a natural experiment, *JMIS* **34** (2017), 457–86.
- [4] M. Maruthappu, A. Hasan, T. Zeltner, Enablers and barriers in implementing integrated care, *Health Syst Reform.* **1** (2015), 250–6.
- [5] I. Köster, I. Schubert, *Epidemiologie und Versorgung von Patienten mit chronischen Wunden. Eine Analyse auf der Basis der Versichertenstichprobe*, AOK Hessen/KV Hessen, 2015. Available from: <https://www.bvmed.de/download/pmv-abschlussbericht-2015-12-03>
- [6] J. Haggerty, R. Reid, G. Freeman, B. Starfield, C. Adair, R. McKendry, Continuity of care: a multidisciplinary review, *BMJ* **327** (2003), 1219–21.
- [7] R. Rudin, D. Bates, Let the left hand know what the right is doing: a vision for care coordination and electronic health records, *J Am Med Inform Assoc.* **21** (2014), 13–6.
- [8] S. Ross, L. Schilling, D. Fernald, A. Davidson, D. West, Health information exchange in small-to- medium sized family medicine practices: motivators, barriers, and potential facilitators of adoption, *Int J Med Inform.* **79** (2010), 123–9.
- [9] J. Gläser, G. Laudel, *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen*, 4th ed., VS Verlag, Wiesbaden, 2010.
- [10] A. Boonstra, M. Broekhuis, Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions, *BMC Health Serv Res.* **10** (2010), 231.
- [11] V. Iacopino, D. Mascia, A. Cicchetti, Professional networks and the alignment of individual perceptions about medical innovation, *Health Care Manage Rev.* **43** (2018), 92–103.

Nursing Record Innovations Aimed at Harmonizing Structured Clinical Knowledge Among Doctors and Nurses

Satoko TSURU^{a,1}, Tetsuro TAMAMOTO^b, Hitoshi FURUYA^b
Akihiro NAKAO^a, Mari FUKUYAMA^b, Kouichi TANIZAKI^a, Naohisa
YAHAGI^a

^a*The University of Tokyo*

^b*Nara Medical University Hospital*

Abstract Hospitalization expenses account for a high proportion of national medical care expenditure in Japan. In 2015, the total national medical care expenditure in Japan was 42.4 trillion yen, and hospitalization expenses were 15.6 trillion yen (36.8%). Therefore, it is necessary to reduce hospitalization expenses. The labor cost of physicians and nurses accounted for about 1/3rd of all expenditure of general hospitals in 2015. Moreover, the personnel cost of nurses accounted for about 1/5th of all expenditure, indicating that it has a marked impact on hospital management. Nurses spend a lot of time completing descriptive records; however, the quality of such records is poor. It is necessary to improve nurse's records to make them highly accessible and reduce the amount of time nurses spend producing records. The objective of this study was to improve the processes underlying record-keeping by nurses in order to harmonize structured clinical knowledge among doctors and nurses. We created 778 Patient Condition Adaptive Path System (PCAPS) items, covering all of the clinical departments that were registered for the PCAPS content master. The resultant masters will be standardized by sharing them with hospitals that adopt the "Team Compass" application. We were able to summarize all of the information in clinical progress sheets because we could link the information described in electronic medical records with that described in Team Compass. Therefore, it became easy to collect information by linking information about clinical orders. The system also made it possible for foundational nursing plans to be created in collaboration with doctors instead of being developed by nurses alone because it allowed information regarding patients' problems, the clinical process, and observation selection to be shared smoothly with doctors. We implemented Team Compass in May 2019. On the first day, PCAPS-based care pathways were used to treat 580 of 623 inpatients. Approximately 4,000 patients were treated using this system from May to August 2019. No major problems have arisen since the implementation of Team Compass.

Keywords: digitalization, quality management, healthcare, structured clinical knowledge

1. Introduction

Hospitalization expenses account for a high proportion of national medical care expenditure in Japan. In 2015, the total national medical care expenditure in Japan was 42.4 trillion yen, and hospitalization expenses were 15.6 trillion yen (36.8%). Therefore, it is necessary to take measures to reduce hospitalization expenses. The total labor cost

of physicians and nurses accounted for about 1/3rd of all expenditure of general hospitals in 2015 [1–3]. Moreover, the personnel cost of nurses accounted for about 1/5th of all expenditure, indicating that it has a marked impact on hospital management.

Nurses spend a lot of time producing records. Specifically, nurses produce descriptive records; however, the quality of such records is poor. Therefore, it is necessary to improve record-keeping systems to make nurse's records easily accessible and reduce the time nurses spend producing records. In a survey evaluating the quality of nursing records, Tsuru et al. suggested that <50% of the patient observations required to deal with individual patients were recorded by nurses. In addition, failures in planning, implementation, and recording have been identified in nursing care in Japan [4–6]. Moreover, it was clarified that data production and input depend on individual nurses, and many observations are recorded in descriptive electronic records, which can be difficult for other staff members to interpret.

To solve planning failures, the preparation and utilization of many high-quality models of treatment and nursing care plans is useful. Thus, we have developed a structured clinical knowledgebase by incorporating the Navigator for Thinking Processes in Nursing (NursingNAVI™) into the Patient Condition Adaptive Path System™ (PCAPS) [7–9].

During this process, structured nursing care plans were developed using the Standard Terminology for Nursing Observation and Action (Ministry of Health, Labour Standards since 2016) for electronic medical records (EMR). Then, an application for supporting the recording of structured nursing care plans was designed and implemented at a large-scale acute-care hospital with 1,000 beds. As a result, the proportion of nurses who took ≥ 30 minutes to identify the conditions of patients in their charge decreased from 45.4% to 5% within 3 months. It was suggested that the utilization of high-quality structured nursing care plans based on clinical nursing knowledge supports the thinking processes of nurses and reduces the time they require to complete their work. After 3 years, the amount of overtime being performed in the hospital had markedly decreased, and almost no overtime was being used to complete records [9].

In Japan, to improve the efficiency of medical care services, the development of an information system to support medical fee payment was initiated in the 1980s, and the development and introduction of a hospital information system (an EMR system) that could record instructions and the execution of tests and treatments was initiated in 2000. Since then, the use of EMR systems has increased, and they have become essential for medical fee payment. In hospitals that use a comprehensive payment system termed DPC, the system records data about any medical care provided, leading to an era of big data in medical care. However, there is a problem with the quality of the data included in EMR, and a survey performed by Yahagi et al. suggested that such records are a long way from being useful for clinical evaluations. It was considered that the development of a new EMR system is necessary to allow analyses of the relationship between hospital management and clinical quality.

Thus, we decided to develop an operational support system that could be used to manage a larger amount of structured clinical knowledge and support routine work.

Since medical care is provided by many departments, 24 hours a day, 365 days a year, it was decided that the system would have to possess the following functions: (1) It should not change existing tests or treatment work and (2) should improve the quality, recording, and execution efficiency of nursing plans.

2. Implementation method

2.1. Development of the ‘Team Compass’ application

We developed ‘Team Compass’ as an application for PCAPS [7–9]. Team Compass has many functions that support the work processes of nurses and doctors. We implemented the first system, which we developed at an acute care hospital with 300 beds, at a university hospital with 900 beds. We listened to the demands of nurses with regard to the functions they wanted and pushed forward with developing these functions. We adopted the following rule to ensure patient safety: Patients’ data should be recorded in both their EMR and the Team Compass system. Therefore, we decided on the following procedure: First, a nurse should enter a patient’s data into their EMR. Then, they should transfer the data to Team Compass.

- 1. A medical doctor orders PCAPS items for each patient from their.
- 2. A nurse opens the EMR system and selects the patient.
- 3. When the nurse clicks the ‘PCAPS’ button, the PCAPS items for the patient are.
- 4. The nurse is able to understand the patient’s condition, and starts their work.

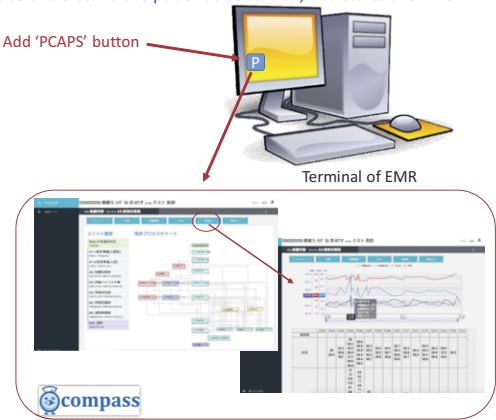


Figure 1. Functions and rules used to ensure patient safety during the use of EMR and Team Compass

2.2. Development of care pathways using PCAPSTM and NursingNAVITM

We interviewed doctors from each clinical department and classified structured treatments into four categories: surgery, internal medicine, short admissions, and one unit. In addition, we linked problems described in the medical task order sheet and Team Compass to share information with doctors, and we constructed a system in which nursing plans are created automatically without checking orders from doctors. We created nursing plans using masters related to surgery and other treatments, which were collaboratively created by doctors and nurses. We used the PCAPS content master developed by the University of Tokyo as the basis for these masters. Also, nursing care and observations completely followed the nursing practice standard masters of the Ministry of Health, Labour and Welfare.

We created 778 PCAPS items, which covered all of the clinical departments that were registered for the PCAPS content master. These masters will be standardized by sharing them with hospitals that adopt Team Compass. We were able to summarize all of the information in patients' clinical progress sheets because we linked the information described in the patients' EMR with that in Team Compass. Therefore, it became easy to collect information by linking information about clinical orders.

The system also made it possible for foundational nursing plans to be created in collaboration with doctors instead of being developed by nurses alone because it allowed information regarding patients' problems, the clinical process, and the selection of observations to be shared smoothly with doctors. In the future, we aim to develop a more sophisticated system that manages the support of patients and reduces the burden of nursing tasks.

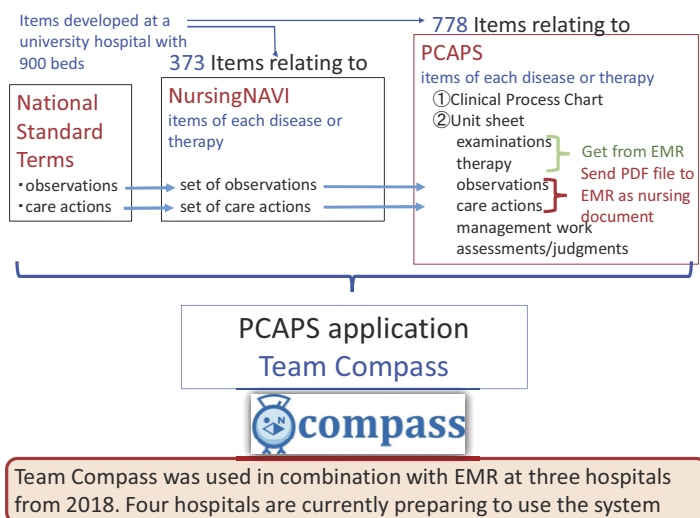


Figure 2. Structured clinical knowledge relating to National Standard Terms for Nursing Practice, NursingNAVI, and PCAPS

3. Results

We implemented Team Compass in May 2019. On the first day, PCAPS-based care was provided to 580 of 623 inpatients. Approximately 4,000 patients were treated using PCAPS from May to August 2019. There were no major problems after the implementation of the system.

The description record of the nurse decreases sharply and, as a structured record, can see it with a progress table. The way of working of the nurse changed. First, they take charge and grasp the summary of the patient from process chart. Secondary, they obtain detailed information with a progress table. The structured record of the nursing is stored within EMR as a PDF file every day in AM at 5:00.

The change of the patient state is detected by a unit shift in the process chart and event occurrence information. The change is shared with a doctor in nurses. The team care between doctor and nurses became in this way smoother.

Nurses handling duties effectively can realize real-time input, and over hours decreases. However, the nurse whom over hours cannot yet shorten exists. We analyze the causes.

4. Discussion

Our system successfully harmonized structured clinical knowledge among medical doctors and nurses. The improvement of record-keeping by nurses at our university hospital has started. We are developing an evaluation system for quality improvement. In the near future, we will implement the evaluation system and advance quality improvement at our hospital.

Acknowledgements

We are deeply grateful for the collaboration of the project members and organizations. This work was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research) No. 17H01608.

References

- [1] Ministry of Health, Labour and Welfare: The general condition of the 2015 national medical expenditure <http://www.mhlw.go.jp/toukei/saikin/hw/k-iryohi/15/dl/data.pdf>
- [2] Summary of the nosocomial business analysis reference index (for 2015 financial statements) <http://www.wam.go.jp/hp/Portals/0/docs/gyoumu/keiei/pdf/shihyou/h27-byouin.pdf>
- [3] Ministry of Health, Labour and Welfare: 2015 hospital management index <http://www.mhlw.go.jp/file/06-Seisakujouhou-10800000-Iseikyoku/0000164133.pdf>
- [4] Satoko TSURU, et.al.: Issues Concerning Volatilizing Situation in Nursing: A Survey on the Nursing Observation for the Post-Gastrectomy Patients, The 14th CJKMI2013, Scientific paper (2013), 33-38p,
- [5] Satoko TSURU, Fumiko WAKO, Miho OMORI and Kumiko Sudo: Problem Solving for Volatilizing Situation in Nursing: Developing Thinking Process Supporting System using NursingNAVI Contents, MIE2015 in Madrid, Scientific paper(2015), 5p,
- [6] Satoko TSURU, Miho OMORI, Manami INOUE and Fumiko WAKO: Quality Evaluation of Nursing Observations based on a Survey of Nursing Documents using NursingNAVI® Contents in JAPAN, The 13th International Congress on Nursing Informatics 2016, Scientific paper (2016), 5pages,
- [7] Satoko Tsuru, Fumiko Wako, Miho Omori, Chitose Watanabe, Mutsuko Nakanishi, Sawako Kawamura: Developing the Structured Knowledge Model to navigate the Nurses' Thinking Process in their Professional Judgment and Action. The 12th International Congress on Nursing Informatics 2014. Scientific paper(2014), 5pages,
- [8] Satoko Tsuru, Yoshinori Iizuka, Masahiko Munechika: Structuring Clinical Nursing Knowledge using PCAPS: Patient Condition Adaptive path System, Proceedings of NI2009(Connecting Health and Humans IOS-Press ISBN 978-1-60750-024-7), scientific paper (2009), 391-395
- [9] Satoko TSURU, Akihiro Nakao, Naohisa Yahagi, Kouichi Tanizaki, Kumiko Sudo, Shizuka Morimatsu, Tomomi Takaki, Nobuko Takakusaki, Tomoko Higashi, Keiko Nakashige, Miyuki Takahashi: Reduction of Overwork Time of Nurses by Innovation of Nursing Records using Structured Clinical Knowledge, MEDINFO 2019: Health and Wellbeing e-Networks for All L. Ohno-Machado and B. Séroussi (Eds.) © 2019 International Medical Informatics Association (IMIA) and IOS Press, 1061-1064, 2019.

Physicians' Experiences of Patient-Initiated Online Consultations in Primary Care Using Direct-To-Consumer Technology

Sabine KOCH^{a,1} and Miwa GUHRES^a

^a*Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden*

Abstract. Both private and public primary healthcare providers increasingly offer their patients online consultation services on request. However, the actual use of these services from a physician's perspective as well as the educational competencies required by the physicians are insufficiently studied. The aim of this study is therefore to explore how general practitioners (GPs) experience video consultations with patients compared to physical consultations in primary care in Sweden. We performed a web-based survey amongst 32 GPs. Despite the advantage of being perceived as time saving, more than half of the physicians did not agree that video consultations are more effective than physical consultations. Most physicians had a positive attitude towards the use of video consultations in their work but reliability of the technical platform was considered to be essential, younger physicians should have worked with physical consultations prior to working with online consultations and the use of (semi-) automatic triage systems was wanted when patients themselves can book appointments for online consultations.

Keywords. Digital visit, online consultation, video consultation, primary care

1. Introduction

Recently the use of mobile technologies related to healthcare delivery has experienced a rapid growth. Direct communication between a healthcare professional and a patient at home or work, so called Direct-to-Consumer (DTC) telemedicine gives patients quick and convenient access to a healthcare professional [1].

In primary care the feasibility and acceptance of utilization of healthcare-initiated telemedicine is supported by many studies and is often more acceptable to patients than healthcare professionals [2]. There are however limited studies about how physicians perceive patient-initiated video consultations. Most studies [3-5] focused on the patients' experience and their satisfaction with using video consultation. Some studies on video consultation investigated physicians' perceptions showing that participants were positive to the use of video consultations in general [6] and elicited potential benefits to use video consultation in primary care [7]. However, at the same time participants felt it should not be used as a substitute for physical consultation and some

¹ Corresponding Author: Prof. Dr. Sabine Koch, Health Informatics Centre, LIME, Karolinska Institutet, Stockholm, Sweden; e-mail: sabine.koch@ki.se

GPs were concerned about the possibility that video consultations would prevent younger physicians from getting clinical experience [7].

In Sweden both public and private care providers have delivered a variety of digital care services and the number of patient-initiated digital visits to physicians is increasing rapidly. Physicians in Sweden have raised concerns regarding their work environment, patient safety and healthcare priorities which have been partly analysed at a smaller scale [8]. Nevertheless, studies of video consultation from a physician's perspective are limited and there is a lack of knowledge about how video consultations are adopted by GPs compared to physical consultations.

Therefore this study aims to explore how general practitioners (GPs) experience video consultations with patients compared to physical consultations in primary care in Sweden.

2. Methods

Data was collected through a web-based survey targeting general practitioners (n=32) in Sweden using exploratory and snowball sampling. The questionnaire consisted of 28 statements to be rated by the GPs on a five-point Likert scale [do not agree at all...completely agree]. The statements were structured according to eight categories; 1. Background, 2. Accessibility, 3. Consultation, 4. Communication, 5. Competence, Education, 6. Care delivery structure, 7. Technology and 8. Open-ended questions regarding the advantages and challenges of video consultation and the competencies needed for physicians in video consultation.

A link to the questionnaire was sent by e-mail to representatives of organizations that offer online video consultations to their patients and to the Swedish Medical Association.

Respondents signed an informed consent when contributing to the study, the data was treated anonymously and was used for research purposes only.

Microsoft Excel and R were used to analyze and visualize quantitative data from the questionnaire. Microsoft Excel and KH Coder were used to analyze and visualize free text/qualitative data.

3. Results

The total number of respondents was 32 with an average age of 42.8 years and the ratio of males to females was 56% and 44%. The average years of experience as GP was 9.6 years. 19 respondents had less than 10 years of experience and 2 respondents had more than 20 years of the experience. The average of years of experience doing video consultation was 1.3 years. The ratio of private, public and both private and public organizations where GPs worked with video consultation was 16%, 78% and 6% respectively. The respondents belonged to 10 different organizations.

28 respondents had access to the patient's health record during video consultations in the same way that they had during physical consultations, four respondents did not. Two third of the respondents (n=22) received education about how to communicate with the patient during video consultations, one third (n=10) did not. The average length of video consultation's time (9.3minutes) was shorter than physical consultation's time (20.8 minutes). Only 1 respondent stated the same length for both

video and physical consultations. All respondents worked with both physical and video consultations, on average 14% of fulltime work was devoted to video consultations.

3.1. Comparison of physical versus video consultations

There was a strong agreement among the respondents that *video consultations give patients faster access to healthcare* [AVG: 4.4; SD: 1.26]², *provide high flexibility to patients* [AVG: 4.3; SD: 0.84], *decrease travel time and costs for patients* [AVG: 4.3; SD: 1.26] and *increase accessibility for persons with functional disabilities* [AVG: 4.1; SD: 1.14]. The respondents were indecisive whether *video consultations increase accessibility for patients with chronic diseases* [AVG: 2.8; SD: 1.70].

The respondents considered *information for decision making during video consultations to be limited compared to physical consultations* [AVG: 4.1; SD: 0.78] but a majority agreed that *video consultations however influence their working environment in a positive way* [AVG: 3.9; SD: 1.27], e.g. through higher flexibility. They were indecisive, however, if *video consultations reduce their workload* [AVG: 3.1; SD: 1.45] and most respondents did not think that *video consultations are more effective than physical consultations* [AVG: 2.2; SD: 1.01]. Also, a slight majority did not *feel comfortable referring patients from a video consultation directly to secondary care* [AVG: 2.8; SD: 1.35].

The respondents were indecisive whether *communication with the patient works as good during a video consultation as during a physical consultation* [AVG: 3.4; SD: 1.27] and whether *GPs need to get extra education about communication and need extra skills when doing video consultations* [AVG: 3.3; SD: 1.13]. The respondents however strongly agreed that *younger physicians need to get clinical experience from physical consultations prior to working with video consultations* [AVG: 4.6; SD: 0.86]. In general, few respondents were *concerned about patients recording the video consultations* [AVG: 1.7; SD: 0.85]. There was not much agreement among the respondents if they *lose the personal contact with the patient during a video consultation* [AVG: 2.8; SD: 1.45].

The respondents further strongly agreed that *a (semi)-automatic triage system needs to be in place* that would guide patients to the right level of care [AVG: 3.9; SD: 1.29] and respondents tended to agree that *it is possible to predict which patient is appropriate for video consultation by analyzing the medical history* [AVG: 3.2; SD: 0.96]. There was also a strong agreement among the respondents that the *technology for video consultations is easy to use* [AVG: 4.2; SD: 0.85] and its *quality (sound, image, reliability of the connection) is good enough* [AVG: 3.9; SD: 0.98].

3.2. Advantages of video consultations

Time efficiency, flexibility regarding working hours and working from home as well as improved working conditions and opportunities for clinical consulting were mentioned as advantages of video consultations.

² Text in this section resembles the statements from the questionnaire that were rated on a five-point Likert scale [do not agree at all...completely agree]. This example: *Video consultations give patients faster access to healthcare* (= Statement); AVG = average Likert value; SD = standard deviation

Time efficiency was supported in the free-text analyses by quotes such as “I get time to meet more patients”, “I meet many patients and can help them urgently” or “Digital visits give a better structure for how the consultation time is used in an efficient way”.

Respondents also pointed out the flexibility of the work specifically by expressing opinions such as “Video consultations give more flexibility for me as physician”, “Flexibility to work from home” and “Flexibility regarding working time”.

They further mentioned advantages regarding their work environment such as “New experience that you would need in case of a general digitalization that you cannot avoid”, “Stimulating and challenging to work in a different way”, “It is very good to be able to follow your own patients”, “Reach patients who are unable to get to the clinic themselves”, and “That patients who are not very sick do not take time slots on the clinics”.

Respondents also discussed video consultations vs phone consultations in their free-text answers exemplifying additional benefits such as “Visual contact is a complement to contact over the phone”, “For example, when dealing with psychiatric disease or skin disease, it is better to have video contact with the patient compared to using phone” and “Gives somewhat more information than a phone call”.

3.3. Challenges of video consultations

The respondents considered it to be challenging to guide the right patients to video consultations, i.e. they should not require a physical examination. This was clearly articulated in free-text answers such as “Need to find the right way in which patients and conditions can best be managed” and “To deselect patients who do not need medical assessment in order to save on healthcare resources”. Further, other forms of digital communication were proposed in some cases - “A lot of consultations can be made with chat not requiring video consultation”. Automatic triage systems were seen as a possible solution to guide patients to the right level of care - “Faster and safer consultations with AI-triage in place”.

Respondents also reported about high expectations and a poor understanding by patients of what can be handled during a video consultation. This is supported by statements such as “Wrong patients are looking for video meetings and they are not those with disabilities but those who are in a hurry in everyday life”, “Some patients have a poor understanding of what can be handled digitally” and “Without triage patients are to a large extent looking for things that are obviously not suitable for video visits”.

Further challenges mentioned by the respondents include the lack of a continuous relationship between physician and patient, the lack of an automatic system for patient triage, the lack of accessibility to integrate with physical care, and the necessity to adapt the current reimbursement model.

4. Discussion and conclusion

The respondents of this study were in general very positive to video consultations. The main advantages mentioned were GPs' flexibility regarding time and working place, increased accessibility for patients and time efficiency for both physicians and patients which is in line with other studies [7-10]. 59% of the respondents were of the

opinion that the lack of a physical examination makes the assessment of the patient insecure during video consultations. This is comparable with the results by Randhawa et al [7] who found that many GPs thought the lack of physical examination was a drawback in video consultations and the assessment of the patient became somewhat incomplete.

The main challenges of video consultations were to find a way to guide the patients to the appropriate consultation that matched with their conditions and the limited information for decision making in video consultations compared to physical consultations. It was considered necessary to introduce a (semi)-automatic triage system before video consultations.

The respondents were indecisive regarding the need to get extra education and need extra skills when doing video consultations. Nevertheless, skills related to digital technology certainly need to be introduced into physicians' educational programs which is also confirmed by a study by Jiwa and Meng [6] who considered video consultation techniques to be required to be learned by medical students in undergraduate education.

In conclusion, most physicians had a positive attitude towards the use of video consultations in their work but reliability of the technical platform was considered to be essential, younger physicians should have worked with physical consultations prior to working with online consultations and the use of (semi-) automatic triage systems was wanted when patients themselves can book appointments for online consultations.

References

- [1] B. M. Welch, J. Harvey, N. S. O'Connell, and J. T. Mc Elligott. Patient preferences for direct-to-consumer telemedicine services: a nationwide survey. *BMC Health Serv. Res* **17** (2017), 784.
- [2] R. L. Bashshur, J. D. Howell, E. A. Krupinski, K. M. Harms, N. Bashshur and C. R. Doarn. The Empirical Foundations of Telemedicine Interventions in Primary Care, *Telemed. J. E. Health* **22** (2016), 342–75.
- [3] R. E. Powell, J. M. Henstenburg, G. Cooper, J. E. Hollander and K. L. Rising, Patient Perceptions of Telehealth Primary Care Video Visits, *Ann. Fam. Med.* **15** (2017), 225–229.
- [4] K. M. McGrail, M. A. Ahuja and C. A. Leaver, Virtual Visits and Patient-Centered Care: Results of a Patient Survey and Observational Study, *J. Med. Internet Res.* **19** (2017), e177.
- [5] S. M. Albert, G. J. Shevchik, S. Paone and G. D. Martich, Internet-based medical visit and diagnosis for common medical problems: experience of first user cohort, *Telemed. J. E. Health* **17** (2011), 304–8.
- [6] M. Jiwa and X. Meng. Video consultation use by Australian general practitioners: video vignette study, *J. Med. Internet Res.* **15** (2013), e117.
- [7] R. S. Randhawa, J. S. Chandan, T. Thomas and S. Singh, An exploration of the attitudes and views of general practitioners on the use of video consultations in a primary healthcare setting: a qualitative pilot study, *Prim. Heal. Care Res. Dev.* **20** (2018), e5.
- [8] J. Alfredsson and S. Koch, Always Open – an analysis of patients' and care professionals' experiences of online appointment bookings and consultations compared to statistics based on EHR data [in Swedish], Health Informatics Centre, Karolinska Institutet, Stockholm, Sweden.
- [9] S. Shaw, J. Wherton, S. Vijayaraghavan et al., Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study, *NIHR Journals Library*, 2018. <https://www.ncbi.nlm.nih.gov/books/NBK507677/>
- [10] R.F. Dixon and J.E. Stahl, A randomized trial of virtual visits in a general medicine practice, *J. Telemed. Telecare* **15** (2009), 115–117.

PM-TOM: A Method for Finding Personalized Polypharmacy Therapies with Minimal Adverse Drug-Drug, Drug-Gene and Drug-Condition Interactions

Adnan KULENOVIC^{a,1} and Azra LAGUMDZIJA-KULENOVIC^a

^a A ! A – Absolute Information Age, Inc. Toronto, Canada.

Abstract. Polypharmacy therapies, quite frequent in older populations, pose significant health risks for patients due to a high possibility of their cumulative adverse drug reactions. We present Personalized Medicine Therapy Optimization Method (PM-TOM) for the discovery of polypharmacy therapies with minimal drug-drug interactions, drug-gene interactions, and drug-condition interactions. The inputs of the PM-TOM heuristic algorithm are the patient's conditions, genome, and the drug products and therapy considered by a clinician. Its output is a comprehensive report that explains details of the optimal therapies for considered drug products and all drugs that treat the patient's conditions. PM-TOM was developed by using the Electronic Medical Records repository of the Personal Genome Project (PGP), and the public repositories: DrugBank and Comprehensive Toxicogenomics Database (CTD). Testing of PM-TOM showed potential for significant reduction of the cumulative adverse drug interactions in personalized polypharmacy therapies. In the group of patients with 8 to 17 conditions, PM-TOM reduced the average cumulative drug interactions from 22 to 5.83, and the group with 6 to 7 conditions from 11.17 to 3. These results encourage further research and development of clinical decision support tools like PM-TOM.

Keywords. Personalized Medicine, Electronic Medical Records, Polypharmacy Therapies, Clinical Decision Support, Adverse Drug Reactions, Drug-Drug Interactions, Drug-Gene Interactions, Drug-Condition Interactions.

1. Introduction

Multiple studies show that polypharmacy therapies can lead to dangerous adverse drug reactions (*ADR*) and high healthcare costs due to their cumulative drug-drug interactions (*DDI*), drug-gene interactions (*DGI*), and contraindicated drug-condition interactions (*DCI*). Paper [1] reports that the average patient in the examined cohort was prescribed with 12 medications; significant *DGIs* were found in 73% of tested patients and *DDIs* in 87% of patients. In the study explained in paper [2], *DDIs* accounted for 66.1% of the total interactions, and the remaining 33.9% were *DGIs*. In Canada, among patients aged 65 years and older with polypharmacy (>5 drugs) admitted to hospital, the prevalence of potential drug interactions is 80% [3]. The study described in the paper [4] states that in 1994 some 2,216,000 hospitalized patients had

¹ Corresponding Author, Adnan Kulenovic, E-mail: A.Kulenovic@rogers.com

severe *ADRs* with 106,000 fatal outcomes, making these reactions between the fourth and sixth leading cause of death in the US that year.

Our research is motivated by the urgency of finding solutions to these troubling facts. In this paper, we present the Personalized Medicine Therapy Optimization Method (*PM-TOM*) for the discovery of polypharmacy therapies with minimal *DDIs*, *DGIs*, and *DCIs* (referred to in the paper as cumulative adverse interactions - *CADIs*). In Section 2, we explain the inputs, outputs, and the database of the *PM-TOM* method, and in Section 3, its heuristic algorithm. In the development, testing, and validation of *PM-TOM*, we used the public data repositories of the *Personal Genome Project* (*PGP*) [5], *DrugBank* [6], and *Comparative Toxicogenomics Database* (*CTD*) [7]. Results, presented in Section 4, show that *PM-TOM* has the potential to aid in the significant reduction of the cumulative drug interactions in polypharmacy therapies.

2. PM-TOM Database, Inputs and Outputs

The components of the ***PM-TOM* database** are (i) Electronic Medical Records (*EMR*) repository (denoted as *EMR*), (ii) Drug Repository (*DRG*), and (iii) Gene-Drug Repository (*GDR*). The current version of *PM-TOM* requires from *EMR* only the patient conditions, drug product prescriptions, and genome (gene variations that could be pathogenic or affect drug behavior, i.e., found to be pharmacogenetic). Other *EMR* elements, such as the patient's gender, race, blood type, drug dosages, etc., will be used in the future *PM-TOM* versions. Formally, $EMR \subset PID \times \mathbb{P}(C) \times \mathbb{P}(G) \times \mathbb{P}(P)$, where *PID* is a set of patient IDs, *C* the set of all conditions, *G* the set of all genes and their variants, *P* set of all commercial drug products, and \mathbb{P} powerset symbol.

The drug repository (*DRG*) includes data about the active drug ingredients (referred to as “drugs” in this paper), commercial drug products (referred to as “drug products”), drug-drug interactions, and drug-condition contraindications. The database keeps drug interactions by referencing active drug ingredients (drugs) rather than drug products. If *D* denotes the set of all drugs, then the set of drug products *P* can be defined as $P \subset D \times N$ (*N* is the set of product names). Set *DC* describes the drug indications, $DC \subset D \times C$. Set $DDI \subset D \times D \times DDS$ defines the drug-drug interactions, where set $DDS = \{1, 2, \dots\}$ quantifies their relative severities. Similarly, set $DCI \subset D \times C \times DCS$ describes the drug contraindications, where set *DCS* quantifies their severities. Gene-Drug Repository (*GDR*) keeps data about drug-gene interactions and their severities, denoted as set $DGI \subset D \times G \times DGS$ (*DGS* is a set of their severities).

***PM-TOM* inputs** include a patient's (i) conditions, (ii) considered drug products, (iii) genome, and (iv) the initial therapy, which is composed of a single drug product for each condition. Formally, the input *PM-TOM* record is $r_i = (pid_i, C_i, G_i, P_i, t_i)$, where $pid_i \in PID$; C_i is the set of the patient's conditions, $C_i \subset C$; G_i is the patient's genome, $G_i \subset G$; P_i is the set of considered drug products, $P_i \subset P$; and t_i is the initial therapy, $t_i \subset C \times P_i$.

***PM-TOM* outputs** are (i) optimal therapy for considered drug products, *tc*, (ii) optimal therapy for all drugs, *to*, which *PM-TOM* creates after examining all drugs from *EMR* applicable to the patient's conditions, and (iii) a detailed personalized therapy report. The report elaborates both optimal therapies, indicating their *CADIs*, and explaining their drug-drug, drug-gene, and drug-condition interactions. It also includes the list of all available drug products for each drug suggested in *to*.

3. PM-TOM Algorithm

In Step 1, *PM-TOM* collects the patient's record. An example is a *PGP* record with patient's *ID* = 'huC29627'; conditions $C_i = \{Anxiety, ADHD, Depression\}$; genome $G_i = \{CBS, TGM1, MBL2, C3, BRCA2, COL41A, TP53, ITPA, KCNH2, DRD2, MBL2, CACNA1C\}$; considered drug products $P_i = \{Venlafaxine, Xanax, Zoloft, Escitalopram, Adderall, Ritalin, Celexa, Wellbutrin, Viibryd\}$; and the initial therapy $t_i = \{(Anxiety, Zoloft), (ADHD, Adderall), (Depression, Celexa)\}$.

PM-TOM identifies, in Step 2, the corresponding drug (drug ingredient) of each considered drug product. For example, the active drug ingredient of *Xanax* is *Alprazolam*. Next, as Step 3, *PM-TOM* prepares the set of pairs of conditions and considered drug products, called CD_i . In the above example, $CD_i = \{(Anxiety, Venlafaxine), (Anxiety, Xanax), (Anxiety, Escitalopram), (Anxiety, Zoloft)^*, (ADHD, Adderall)^*, (ADHD, Ritalin), (Depression, Celexa)^*, (Depression, Viibryd), (Depression, Wellbutrin)\}$. The pairs marked with * indicate the initial therapy (t_i).

In Step 4, *PM-TOM* uses set CD_i to identify candidate therapies, t_k , for considered drug products. It forms candidate therapies by replacing only one drug in the initial therapy with another drug from set CD_i applicable to the same condition. An example of a candidate therapy is $\{(Anxiety, Xanax), (ADHD, Adderall)^*, (Depression, Celexa)^*\}$, i.e., for *Anxiety*, *Xanax* replaced *Zoloft*. Note that *PM-TOM* does not evaluate all combinations of considered drug products so that it could miss the most effective therapy. However, the testing results show that this approach can lead to significant improvements in the initial treatments. If d is the average number of candidate drugs for n conditions, then the complexity of an algorithm that evaluates all drug combinations will be proportional to d^n . *PM-TOM* examines only $n*d$ candidate therapies.

In Step 5, *PM-TOM* calculates *CADI* indicator (w) of each candidate therapy t_k after examining the following sets: (i) drug-drug interactions $DDI_k = \{ddi_k = (d_1, d_2, dds): \exists(c_1, d_1) \in t_k, \exists(c_2, d_2) \in t_k, ddi_k \in DDI\}$, (ii) drug-gene interactions $DGI_k = \{dgi_k = (d, g, dgs): \exists(c, d) \in t_k, g \in G_i, dgi_k \in DGI\}$, and (iii) drug-condition contraindications $DCI_k = \{dci_k = (d, c, dcs): \exists(c_1, d) \in t_k, \exists(c, d_1) \in t_k, \nexists(c_1, d_1) \in t_k, dci_k \in DCI\}$. $w(t_k)$ is calculated as

$$w(t_k) = DDIP * \sum_{ddi_k \in DDI_k} ddi_k[dds] + DGIP * \sum_{dgi_k \in DGI_k} dgi_k[dgs] + DCIP * \sum_{dci_k \in DCI_k} dci_k[dcs]. \quad (1)$$

DDIP, *DGIP*, and *DDIP* are the *PM-TOM* parameters that define relative weights of *DDIs*, *DGIs*, and *DCIs* (set to 1 by default). $ddi_k[dds]$, $dgi_k[dgs]$, and $dci_k[dcs]$ indicate severities of these interactions. As an example, the *CADI* indicator of the above initial therapy is $w(t_i) = 6$.

Next, in Step 6, the algorithm iterates through all conditions, and for each condition, finds candidate therapy with minimal *CADI*. In Step 7, *PM-TOM* uses the drugs for each condition that form these candidate therapies to create the optimal therapy for considered drug products. In the above case, the optimal therapy is $t_c = \{(Anxiety, Venlafaxine)^{\#}, (ADHD, Ritalin)^{\#}, (Depression, Celexa)^{\#}\}$, with $w(t_c) = 4$.

As mentioned above, *PM-TOM* is an iterative method. It repeats the above steps, taking the optimal therapy found above as the initial therapy for the next iteration. If the new candidate optimal therapy shows no better $w(t_c)$, the algorithm will stop (Step 8).

In the next phase, *PM-TOM* repeats Steps 3-8 to find the optimal therapy for all drugs (t_o) after (i) collecting all drugs from the *EMR* used to treat the patient's conditions and (ii) taking already found therapy t_c as the starting point, instead of therapy t_i . In the above example, *EMR* returned 33 applicable drugs. t_o was found as $\{(Anxiety, Butabarbital)^{\#}, (ADHD, Dexmethylphenidate\ HCL)^{\#}, (Depression, SAM-E)^{\#}\}$, with $w(t_o)=1$. The report for this *EMR* can be found at <https://www.abs-info-age.com/project-1>.

4. PM-TOM Validation and Testing

In the development, testing, and validation of *PM-TOM*, we used the *EMRs* of *PGP*, which include the patient conditions, genome, and administered drug products. *DrugBank* provided the drug product, drug-condition, and drug-drug interaction data, and *CTD* the drug-gene interaction data. We tested 286 *EMRs*. Table 1 lists in column (A) various categories of tested *EMRs*, which have the same number of conditions. Column (B) shows the count of cases and (C) the average number of genes in each category. Columns (D) and (E) display the average number of considered drug products and all drugs applicable to the patient's conditions.

Table 1. Average *CADI* of the optimal therapies for the considered drugs and all applicable drugs.

(A) Conditions	(B) Cases	(C) Genes	(D) Consid. drugs	(E) All drugs	(F) <i>CADI</i> - Consid. drugs	(G) <i>CADI</i> - All drugs
1	77	13.58	1.31	6.12	0.6	0.09
2	119	5.74	2.34	10.6	1.05	0.22
3	49	3.43	3.51	16.71	2.1	0.33
4	16	4.31	4.38	20.88	3.38	0.56
5	13	4.46	5.92	26	4.92	2.08
6..7	6	2.33	7.33	34	11.17	3
8-	6	10	12.83	42	22	5.83

Columns (F) and (G) display the average *CADIs* of the optimal therapies for considered drug products and all applicable drugs, which indicate a significant reduction of the average *CADI* between these two therapies (approximately three to six times, depending on the category). The algorithm required two iterations (Figure 1).

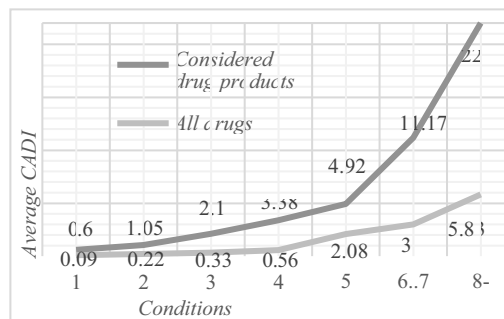


Figure 1. Average *CADI* for optimal therapies

5. Discussion

In paper [8], the author underlines the need for highly personalized drug selection by expert systems. To the best of our knowledge, *PM-TOM* is an original approach to do so. The testing results presented in Section 4 show that the method can identify possible treatments with significantly decreased *CADIs*. We also tested *EMRs* for some recurrent multi-diseases, like *{Hypertension, Diabetes Mellitus, Hyperlipidemia}*, and their *CADI* reduction was approximately five times. A data mining algorithm for finding frequent multi-diseases, and their possible monogenic causes, is explained in paper [9].

Why do prescribed therapies typically have a high number of drug interactions? We speculate that the reasons are the lack of the patient's genetic data, which would enable finding highly personalized therapies, and the limitations of used tools, which are typically focused on *DDI* alerts. Also, according to the investigation in the US, reported in paper [10], these alerts are overridden for various reasons in about 90% of cases.

PM-TOM is an efficient iterative heuristic algorithm. Its testing confirmed that very few iterations lead to finding therapies with significantly decreased *CADIs*. The algorithm complexity is rather low as it evaluates a small subset of candidate therapies. For example, for seven candidate drugs for 12 conditions, *PM-TOM* will examine only 84 candidate therapies per iteration of possible 14 billion ones. Due to that feature, *PM-TOM* has room for extension with other features relevant for polypharmacy therapies, such as (i) rule-based medical guidelines, (ii) lab test results, (iii) drug dosages, (iv) patient's demographic data (age, race, gender, environment), (v) drug-drug-gene interactions (*DDGIs*), (vi) drug-food interactions, (vii) drug-microbiome interactions, and others.

References

- [1] K Reynolds, D. Pierce, F. Weitendorf, M. Linder: Avoidable drug-gene conflicts and polypharmacy interactions in patients participating in a personalized medicine program. *Personalized Medicine* **14** (2017), 221-233.
- [2] P. Verbeurgt et al., How common are drug and gene interactions? Prevalence in a sample of 1143 patients with CYP2C9, CYP2C19, and CYP2D6 genotyping, *Pharmacogenomics*. **5** (2014), 655-65.
- [3] C. Tannenbaum, N. Sheehan, Understanding and preventing drug-drug and drug-gene interactions, *Expert Reviews, Clinical Pharmacology*, **7** (2014), 533-544.
- [4] J. Lazarou, B. Pomeranz, P. Corey, Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies, *JAMA* **15** (1998), 1200-1205
- [5] M. Ball et al., Harvard Personal Genome Project: Lessons from participatory public research, *Genome Medicine* **6** (2014).
- [6] D. Wishart, Y. Feunang A. Guo, et al., DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Research* (2018), D1074-D1082.
- [7] A. Davis, C. Grondin, R. Johnson, D. Sciaky, R. McMoran, J. Wiegiers et al., The Comparative Toxicogenomics Database: update 2019, *Nucleic Acids Research*. (2019), D948-D954.
- [8] C. Hizel. Highly personalized reports for personalized drug selection by expert systems as clinical decision support, *Personalized Medicine* **14** (2017), 93-97.
- [9] A. Kulenovic, A. Lagumdzija-Kulenovic: Discovering Monogenic Causes of Multi-Diseases by Mining Electronic Medical Records and Genetics Repositories, *Studies in Health Technology and Informatics* **258** (2019), 189-193.
- [10] A. Wright et al., Structured override reasons for drug-drug interaction alerts in electronic health records, *Journal of the American Medical Informatics Association* **26** (2019), 934-942.

Practical Implementation of Receiver-Oriented Encryption in STROKE OWL

Timo MICHELSEN^{a,1}, Christian LINS^a, Andreas HEIN^b, Christian LÜPKES^a

^a*OFFIS – Institute for Information Technology, Div. Health,
Escherweg 2, 26121 Oldenburg, Germany*

^b*Carl von Ossietzky University Oldenburg, Dep. Health Services Research,
Ammerländer Heerstr. 140, 26129 Oldenburg, Germany*

Abstract. New forms of care for complex chronic diseases require substantial efforts in the collection, storage, and analysis of privacy sensible medical data. Additionally, providing practical support for those who coordinate the actual care management process within a diversified network of regional service providers is beneficial. In this paper, we present a follow-up with the current status, lessons learned, and preliminary results of the IT infrastructure in project STROKE OWL. The project aims at a comprehensive implementation of cross-sector care management for stroke patients. Patients are accompanied in their recovery by stroke pilots, who use a specially designed tablet app for the piloting. It is shown how the tablet app uses receiver-oriented encryption to support the piloting process while ensuring privacy.

Keywords. STROKE OWL, Data Management, Privacy by Design, ROE

1. Introduction

The project STROKE OWL [1] (located in the pilot region Ostwestfalen-Lippe, Germany) aims at a comprehensive implementation of cross-sector care management for patients after a stroke. The actual care management process is carried out by so-called *stroke pilots*, who accompany patients in coordinating their lives after the stroke incident like case managers. In a population-oriented approach, the complete care path is integrated, i.e. the formation and coordination of a network of regional service providers consisting of stroke units, rehabilitation partners, and ambulatory actors, as well as health insurance funds. The primary goal of the project is to measure the change of recurrence rates due to this new care management process.

The mere number and diversity of individual actors (stroke pilots, health insurance funds, study nurses etc.) place high demands on the underlying IT infrastructure of such projects about the collection, provision, and evaluation of data, especially concerning security and privacy. Due to the overall demand for Privacy by Design [2], we proposed a model of comprehensive receiver-oriented encryption (ROE) [3], which we are using in STROKE OWL. In this paper, we show how ROE is practically implemented in the accompanying LotsenApp, a tablet application that supports the stroke pilots in their work with the patients, and how ROE is used in the central data structure. We also present

¹Corresponding Author, email. timo.michelsen@offis.de

a follow-up with an update about the current status, lessons learned with ROE and its implementation in STROKE OWL.

In the following Section 2, we give a brief overview of ROE. Section 3 describes the implementation details of the LotsenApp tablet app and our system architecture motivated by ROE. In Section 4, we give a small overview about some lessons learned during this development. Finally, we conclude with a small status update of STROKE OWL with over 1000 patients and 1 year of app usage.

2. Receiver-oriented encryption (ROE) in STROKE OWL

ROE [3] aims to restrict the access of data recipients using encryption so that only they can access their data and only their data. Especially for medical data it is essential that their accessibility is as limited as possible. To enforce this ROE combines and adapts two existing methods: asymmetric cryptography [4] and pseudonymization.

In STROKE OWL, we realized ROE as follows [3]: stroke pilots gather needed primary data of their accompanied patients. *Primary data* is a combination of personal data (name, address, etc.), domain-specific data (e.g., cross-sectoral medical data from doctors), and case management data for the stroke pilots (e.g., case schedules, assessments, contact network management). The collected primary data is then encrypted with the specific recipient key and stored at the central servers (as *data packets*). Due to asymmetric encryption, the server infrastructure operators are not able to decrypt them. Since each recipient gets their encrypted data, it was necessary to create individual private-public-key pairs for asymmetric encryption.

The following recipients have warranted demands on these packets: two study nurses (to send paper-based questionnaires to the patients), seven health insurance funds (to provide additional *secondary data*), and the University of Bielefeld (UBF) (which gets secondary data, parts of primary data, and questionnaires answers to finally evaluate STROKE OWL). Each of these recipients gets their own data packets.

In ROE, the server-application sends the data packets to their respective recipients for decryption (only they have the needed individual private keys). To avoid the possibility to merge the decrypted packages again (e.g., to re-identify a patient), we proposed in ROE to send a different pseudonym (PSN) for each recipient (so-called *recipient-oriented PSNs*, *ROPSN*). The server-application uses a central PSN for each patient to (1) store the encrypted data packages in a consistent way and (2) for *identity management* - to translate between different ROPSN when needed.

3. Development and system architecture

With the early focus on Privacy by Design, our system design process was heavily driven by data security and safety. Especially the first implementation and later updates for ROE forced us to document and evaluate every single design decision. This led us to two important documents: (1) a detailed description of needed data attributes, formats, transfer protocols, roles and role interaction processes (*data format and transmission concept*) and (2) a *data security and privacy concept*, which describes measures to secure this data in the context of STROKE OWL. After all project partners agreed to these

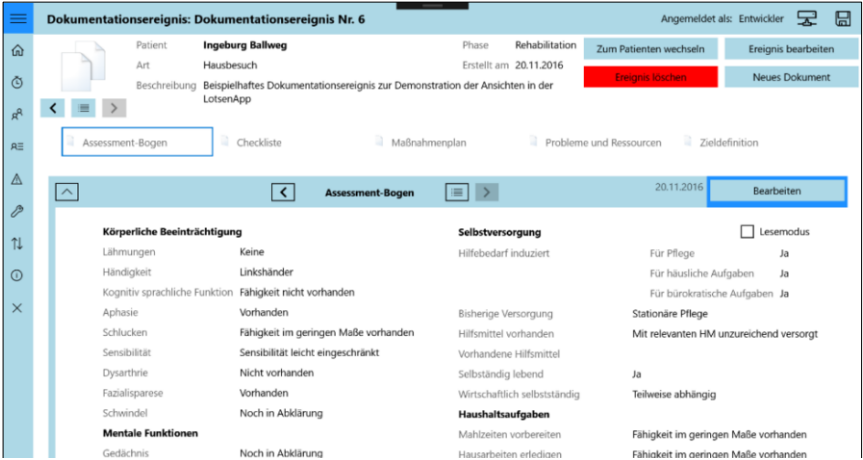


Figure 1: Example screenshot of the LotsenApp (with generated example primary data)

documents and concepts, we designed and implemented our IT system architecture accordingly with ROE in mind (see Figure 2).

Our current architecture is divided into three parts: (1) client application, (2) server application and (3) recipient application(s). The client application *LotsenApp* is used by each stroke pilot (see an example screenshot in Figure 1). It is a graphical user interface which helps the pilots to input primary data fast and easy. We defined well-formed selection lists, added simple plausibility checks, and developed a documentation structure to manage the huge number of different fields needed to fill for each patient. Additionally, the *LotsenApp* contains tools to manage the contact networks of patients, doctors etc. easily.

The *LotsenApp* is the only application which has a reason to access, edit, and delete all primary data. It encrypts (parts of) this data asymmetrically with the individual public keys of all recipients. If the client application has access to the internet, encrypted data packets are sent to the server application.

The *LotsenTablet* is a mobile tablet the *LotsenApp* is running on and allows stroke pilots to access the data at the patients side everywhere (in hospitals, at patients, etc.).

The realization of the server-application is called *DAVE* (Data Vault Environment). *DAVE* (1) manages and distributes public keys for data encryption, (2) stores encrypted data packages for each recipient, and (3) contains the identity management for translating and transmitting PSNs and ROPSNs.

Since the collected primary data in STROKE OWL is essential for the success of this project, we also implemented a three-layered backup system to minimize the possibility of data loss or unintended changes. This procedure is independent to ROE. In the first layer, the *LotsenApp* stores primary data always locally at first (in so called *save states*). Every save state creates a new local file (up to 200 files). With this, stroke pilots can “jump back” to older save states to recover from errors or data loss. If a connection to *DAVE* is available, the *LotsenApp* transmits a backup copy to *DAVE* (second layer to avoid data loss due to hardware failures). Finally, in the third layer, *DAVE* creates backups of all its databases.

To maintain data privacy, all backup copies are symmetrically encrypted with a *data password*, chosen and only known by the specific stroke pilot. Data passwords are also

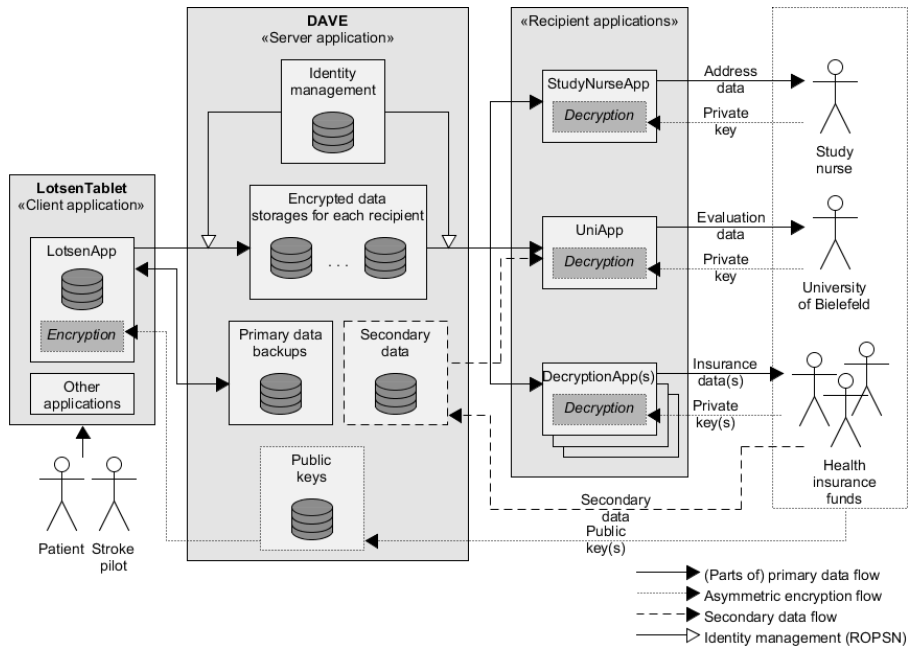


Figure 2: Overview of the IT system architecture in STROKE OWL

independent from our authentication mechanisms. Therefore, each stroke pilot has two passwords: one for authentication and one for their primary data.

Instead of using one central server application for implementing DAVE, we decided to move towards a modular design with multiple, smaller, and independent server applications (connected with REST-APIs and HTTPS). Some of them are the following: primary data backup, authentication and authorization management, identity management (ROE), secondary data management, recipient data package storage (ROE), and data import/export. There is also a mirrored (but separate) server of DAVE (called *staging server*), which allows us to test data loss, connection problems, new developments, changes, and corrections in more realistic environments (integration and system tests). This avoids problems of losing a tablet or a hacking attempt stated in [5].

The recipient applications proposed in ROE are represented by multiple smaller applications in STROKE OWL. In each recipient application, the user (study nurse, UBF and health insurance funds) provides his private key to decrypt the data packages, which were sent from DAVE beforehand. The *StudyNurseApp*, the recipient application for study nurses in STROKE OWL, additionally helps in coordinating the paper-based questionnaires (marks, which questionnaires were sent, received, etc.).

4. Lessons learned

Defining and communicating the data format and transmission concept as well as the data security and privacy concept played a key role in implementing very important security features like asymmetric encryption and ROE. Both documents helped us to communicate with our project partners in technical, organizational, and computational concerns. Especially health insurance funds, who will provide secondary data, needed these descriptions. In similar (health-specific) projects, we strongly encourage to define

these documents beforehand to avoid security-related problems and misunderstandings between different sectors, persons and stakeholders.

Additionally, the three-layered backup system to avoid data loss is beneficial. Despite the fact that backups are necessary measures for data safety, three backup layers also offer a remedy against user-induced, non-grave data loss. Therefore, not every data loss leads to a complete recovery of the entire IT-infrastructure saving time and money.

The generation of receiver-specific key-pairs was more time-consuming and organisationally complicated as expected. For similar projects with many different recipient, we need to dedicate more time for this crucial step. In STROKE OWL, we compensated the time-loss with a developed simple application which creates key pairs automatically. Every recipient created a pair in one minute. For decryption, we provide another application. With this, recipients can use their primary keys to decrypt their encrypted data.

5. Current state and conclusion

We outlined the practical implementation of receiver-oriented encryption on the basis of the STROKE OWL project. Privacy by Design is a complex but important design paradigm especially in health-care. In STROKE OWL, this resulted in our proposed ROE. ROE can be used in similar projects with high data security demands and many data recipients with focus on very strict asymmetric encryption rules. However, due to its organisational complexity with that many project partners, a long planning phase must be considered to avoid later problems. The success of STROKE OWL shows us that this drawback is worth to take: As of January 2020, currently 17 stroke pilots are using the LotsenApp successfully. More than 1200 patients are participating in this project. Two study nurses using the StudyNurseApp. Up to 3200 paper-based questionnaires were sent and filled successfully. The reports from pilots and other users showed high satisfaction and productivity with the applications. We are content with this development and continuously refine and update our applications and components.

Acknowledgments

This work was funded by the Federal Joint Committee German Innovation Fund within the joint research project STROKE OWL (grant no. 01NVF17025).

References

- [1] Stiftung Deutscher Schlaganfallhilfe (SDSH), STROKE OWL Project Website, <https://stroke-owl.de/de/startseite/> (last visit at 14th October 2019).
- [2] Schaar, P., Privacy by Design, *Identity in the Information Society*, 3(2) (2010), 267–274.
- [3] Michelsen, T., Lins, C., Gudenkauf, S., Hein, A., Lüpkes, C., Privacy by Design for Integrated Case and Care Management: Receiver-oriented encryption in STROKE OWL, *Studies in Health Technology and Informatics* 258 (2019), 110–113. doi: 10.3233/978-1-61499-959-1-110
- [4] Benaloh, J., Chase, M., Horvitz, E., Lauter, K., Patient controlled encryption: ensuring privacy of electronic medical records, *Proceedings of the 2009 ACM workshop on Cloud computing security* (2009), 103–114.
- [5] Miller, A.R., Tucker, C.E. Encryption and the loss of patient data. *J. Pol. Anal. Manage.*, 30 (2011); 534–556. doi:10.1002/pam.20590

Relationship Between Exercise Duration in Multimodal Telerehabilitation and Quality of Sleep in Patients with Multiple Sclerosis

In cheol JEONG^a, Herbert KARPATKIN^b, Joel STEIN^c, Joseph FINKELSTEIN^{a,1}

^a*Icahn School of Medicine at Mount Sinai*

^b*City University of New York*

^c*Columbia University, New York, NY, USA*

Abstract. The purpose of this study was to investigate the effect of a telerehabilitation system on the quality of sleep in patients with multiple sclerosis (PwMS). Fifteen females and two males (60.1 ± 11.4 years) who used the system for three months completed the Pittsburgh Sleep Quality Index (PSQI) at the baseline and end of follow-up. Total System Usage (TSU) and Total Exercise Time (TET) were elucidated from the system web logs for each PwMS. A significant association ($p < 0.05$) was found between PSQI sleep efficiency (SE) and TSU (0.76) and between SE and TET (0.81). The association between PSQI total score (TS) and TSU and between TS and TET were -0.507 and -0.702 respectively ($p < 0.05$). Our results uncovered an association between amount of exercise time spent by PwMS and positive effects on both the efficiency and quality of sleep. Thus, further development of approaches promoting continuous participation of PwMS in telerehabilitation is warranted.

Keywords. Telerehabilitation, exercise duration, multiple sclerosis

1. Introduction

Sleep disturbances in people with multiple sclerosis (PwMS) are associated with fatigue, depression, and reduced quality of life [1]. Over 60% of PwMS report impaired sleep quality [2]. In middle-aged and older adults with sleep problems, an exercise program improved the quality of sleep [3]. Current evidence indicates that patients with multiple sclerosis can gain health benefits similar to those from healthy populations through regular exercise [4, 5]. Home-based telerehabilitation has been shown to be well accepted by PwMS [6, 7]. The Home Automated Telemanagement (HAT) system supporting regular multimodal exercise was described previously as promising means to improve mobility and improve quality of life in PwMS [8]. The sustainable impact of remote disease management depends on patient adherence to the program [9]. In addition, high levels of participation, which include proper exercise engagement, facilitate improvements in patients' quality of life, curb the negative effects of fatigue, and improve the quality of sleep. However, the association between extent of

¹ Corresponding Author, Joseph FINKELSTEIN, MD, PhD, Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; E-mail: joseph.finkelstein@mssm.edu.

participation in home-based exercise promulgated by a telerehabilitation system and changes in sleep quality has not been systematically evaluated. The main purpose of this study was to investigate the effects of a telerehabilitation system usage patterns on the quality of sleep in PwMS.

2. Methods

2.1. Systems

The HAT system supporting individualized home-based telerehabilitation in PwMS was used in this study [6]. Following the chronic care model, the HAT system guides PwMS in executing their individualized exercise plans at home, facilitates patient-provider communication and multidisciplinary care coordination as previously described [8]. Tailored motivational messages are provided to enhance adherence [6-8].

2.2. Study design

After being consented as a study participant, PwMS were prescribed an individualized exercise plan based on a comprehensive physical therapy examination by a licensed physical therapist. All patients were trained how to perform exercise at home by the physical therapist and received the necessary exercise equipment. The study participants were asked to use the home-based telerehabilitation system on a daily basis. The system guided them in following their exercise prescription, logged their exercise performance, and communicated exercise logs to their physical therapist who followed up with the patients on as needed basis based on patient performance.

The Pittsburg Sleep Quality Index (PSQI) [10] was used to evaluate self-reported quality of sleep at baseline (BL) and 3-month follow-up (3M). Sleep quality was characterized by four PSQI factors: sleep efficiency, sleep disturbances, daytime dysfunctions, and total score.

The patterns of telerehabilitation system usage including extent of regular exercise completion over the 3-month period were elucidated based on daily web-based logs. The web logs were automatically generated by the HAT system which included time stamps for each web page visited by the telerehabilitation system users. Based on these logs, the following telerehabilitation usage indicators were calculated:

- Total system usage time: total system usage time over 3-month period from BL to 3M
- Daily usage time: average time spent with the system on the days when a user logged in (users could log in multiple times during a single day)
- Visit usage time: average time spent with the system for each separate log-in instance
- Average percent of sets completed: average percent of exercise sets completed out of all sets prescribed for a daily exercise program
- Total exercise time: total time spent in an exercise mode in the telerehabilitation during the 3-month period from BL to 3M

The system usage indicators provided insights on how different patients used the telerehabilitation system and how much time they devoted to the prescribed exercise

program. Associations between the PSQI scores and system usage indicators were analyzed using IBM SPSS Statistics 26 for Windows. Association between study variables (r) was assessed using Pearson correlation coefficient.

3. Results

3.1. Socio-demographic characteristics of patients

Overall, 3-month telerehabilitation data from seventeen PwMS were analyzed (15 females and 2 males). The average patient age was 60.1 ± 11.4 ranging from 39 to 76 years old; 70.6% of the participants were Whites and 11.8% were Hispanics; average duration of education was 15.5 ± 2.7 years. The time since multiple sclerosis diagnosis varied between 2.5 and 50.8 years and the average of education was. Self-assessed disease severity was reported as severe by 11.2% of participants; 29.4% reported that their disease is progressing (Table 1).

Table 1. Socio-demographic Characteristics of Patients

Characteristics (yrs.)	Minimum	Maximum	Mean	SD
Age	39	76	60.1	11.4
How long have MS	2.5	50.8	26.0	13.5
School	10	20	15.5	2.7
Characteristics	Gender	Race	Hispanic	
No. (%) (N=17)	female	AA	3(17.6)	yes
		Asian	2(11.8)	
	Male	white	12(70.6)	no
				15(88.2)
Characteristics	MS severity	Disease		
No. (%) (N=17)	Mild	5(29.4)	progressing	5(29.4)
	moderate	10(59.8)	stable	11(94.1)
	severe	2(11.2)	improving	1(5.9)

yrs.: years, MS: multiple sclerosis, SD: standard deviation, AA: African American

3.2. PSQI and system usage patterns

Average changes between BL and 3M for PSQI sleep efficiency, PSQI sleep disturbances, PSQI daytime disturbances, and PSQI total score were -0.1, 0.9, 0.5 and 0.2 respectively (Table 2).

The total time that PwMS spent using the HAT system over the 3-month follow-up period ranged between 6 and 68 hours with average time of 30.7 ± 19.8 hours. The mean total time patients spent with a system on a day when they logged-in was 26.5 minutes, and the average total time patient spent at each single sign-in visit was 22.0 minutes. The patients' average percentage of completed exercise sets and prescribed number of exercise repetitions were 46.5% and 72.6% respectively, and the average total time spent in the exercise mode was 13.5 hours (Table 2).

3.3. Bivariate correlations

Pair wise associations between nine sleep quality and system usage parameters were investigated including 4 PSQI scores and 5 usage pattern indicators. The association between PSQI sleep efficiency and Total system usage time and between PSQI sleep

efficiency and Total exercise time were 0.757 and 0.814 respectively. The association between PSQI total and Total system usage time and between PSQI total and Total exercise time were -0.507 and -0.702 respectively. Correlation between Daily usage time and Visit usage time was 0.910 and r between Total system usage time and Total exercise time was 0.886. The r between PSQI sleep efficiency and PSQI total was 0.722.

Table 2. PSQI scores, PSQI changes in 3 months, and system usage indicators

Parameter name	Minimum	Maximum	Mean	SD
PSQI sleep efficiency (BL)	57	100	84.7	12.6
PSQI sleep efficiency (3M)	56	120	84.5	16.2
PSQI sleep efficiency (3M-BL)	-23.3	38.3	-0.1	15.9
PSQI sleep disturbances (BL)	1	11	6.4	3.3
PSQI sleep disturbances (3M)	3	18	7.3	4.1
PSQI sleep disturbances (3M-BL)	-6	8	0.9	4.6
PSQI daytime dysfunctions (BL)	0	3	1.0	1.0
PSQI daytime dysfunctions (3M)	0	4	1.5	1.4
PSQI daytime dysfunctions (3M-BL)	-1	3	0.5	1.1
PSQI total (BL)	1	11	6.7	2.8
PSQI total (3M)	2	14	6.9	3.4
PSQI total (3M-BL)	-6	3	0.2	2.3
Total system usage time (sec)	21,494	244,042	110,569.9	71,212.8
Daily usage time (sec)	311.5	4,258.3	1,588.2	947.9
Visit usage time (sec)	307.1	2,248.8	1,324.8	601.0
Average percent set completed (%)	69.4	142.0	97.8	17.7
Total exercise time (sec)	7,489	157,502	48,457.2	45,499.8

SD: standard deviation, BL: data acquired at the baseline, 3M: data acquired at the 3-month follow-up, 3M-BL: difference in PSQI score between 3M and BL

Table 3. Correlation matrix between PSQI score changes and system usage indicators

Parameter name								
A. PSQI sleep efficiency (3M-BL)	A							
B. PSQI sleep disturbances (3M-BL)	-0.041	B						
C. PSQI daytime dysfunctions (3M-BL)	0.562	0.362	C					
D. PSQI total (3M-BL)	-0.722	0.248	-0.105	D				
E. Total system usage time (sec)	0.757	0.085	0.400	-0.507	E			
F. Daily usage time (sec)	0.536	-0.109	0.490	-0.240	0.588	F		
G. Visit usage time (sec)	0.532	-0.153	0.306	-0.302	0.596	0.910	G	
H. Average percent set completed (%)	0.102	-0.255	-0.047	0.166	-0.098	0.135	0.278	H
I. Total exercise time (sec)	0.814	0.127	0.321	-0.702	0.886	0.369	0.436	-0.114

All values represent the Pearson correlation coefficient (r), light gray cells have p value<0.05 and dark gray cells have p-value<0.01, BL: data acquired at baseline, 3M: data acquired at the 3-month follow-up, 3M-BL: score difference between 3M and BL

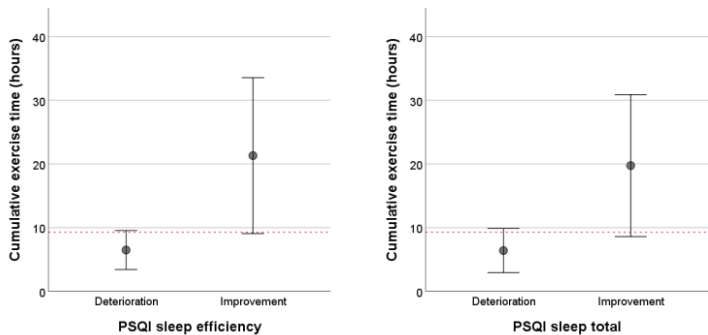


Figure 1. Comparisons between the enhanced group and the degenerated group

Table 4. Comparison between sleep improvement and deterioration groups (mean ± standard deviation)

Group	TSU	TET	ΔSE	ΔTS
Sleep Improvement	38.6±20.9	19.7±14.5	12.7±12.7	-1.4±2.0
Sleep Deterioration	22.0±15.2	6.4±4.1	-11.6±7.3	2.0±0.8

TSU: total system usage in 3 month (hours), TET: total exercise time in 3 month (hours),
ΔSE: change in PSQI sleep efficiency score, ΔTS: change in PSQI total score

4. Discussion

In this study, we used the a telerehabilitation system supporting an individualized multimodal exercise prescription to investigate how system usage patterns are associated with sleep quality. Our results indicate the accumulation of the system usage time has a positive effect on both the efficiency and quality of sleep. In particular, the accumulation of exercise time had a greater impact on PSQI scores (Table 4). We found that there was a significant difference in the total exercise time between the PwMS with improved or unchanged sleep efficiency and those whose sleep efficiency deteriorated ($F=8.570$, $p=0.010$). Similarly, there was a significant difference in the total exercise time between the group of patients with improved or unchanged total quality of sleep (PSQI total) and the group of patients whose total sleep quality score deteriorated over 3-month follow-up period ($F=6.247$, $p=0.024$) as depicted in Figure 1.

Based on the results from 3-month telerehabilitation, improvements in sleep quality were associated with approximately 9.3 hours of total exercise time which is an equivalent of at least 20 minutes of exercise twice a week. Our data strongly support positive impact of telerehabilitation on PwMS who might otherwise have difficulty traveling to an outpatient program. Further work to enhance PwMS participation in telerehabilitation is warranted.

References

[1] J.P. Zajicek, J.C. Hobart, A. Slade, D. Barnes, P.G. Mattison, MUSEC Research Group, Multiple sclerosis and extract of cannabis: results of the MUSEC trial, *Journal of Neurology, Neurosurgery & Psychiatry* **83** (2012), 1125-1132.

[2] B.R. Stanton, F. Barnes, E. Silber, Sleep and fatigue in multiple sclerosis, *Multiple Sclerosis* **12** (2006), 481-486.

[3] P.Y. Yang, K.H. Ho, H.C. Chen, M.Y. Chien, Exercise training improves sleep quality in middle-aged and older adults with sleep problems: a systematic review, *Journal of Physiotherapy* **58** (2012), 157-163.

[4] J.H. Petajan, A.T. White, Recommendations for physical activity in patients with multiple sclerosis, *Sports Medicine* **27** (1999):179-191.

[5] J.N. Slawta, A.R. Wilcox, J.A. McCubbin, D.J. Nalle, S.D. Fox, G. Anderson, Health behaviors, body composition, and coronary heart disease risk in women with multiple sclerosis, *Archives of Physical Medicine and Rehabilitation* **84** (2003), 1823-1830.

[6] J. Finkelstein, J. Liu, Usability of telerehabilitation system supporting multipronged exercise in patients with multiple sclerosis, *Stud Health Technol Inform* **251** (2018), 281-284.

[7] I.C. Jeong, J. Finkelstein, Introducing telerehabilitation in patients with multiple sclerosis with significant mobility disability: pilot feasibility study, *IEEE ICHI* (2015), 69-75.

[8] J. Finkelstein, J. Wood, E. Cha, Impact of physical telerehabilitation on functional outcomes in seniors with mobility limitations, *IEEE EMBS* (2012), 5827-5832.

[9] I.C. Jeong, J. Liu, J. Finkelstein, Factors Affecting Adherence with Telerehabilitation in Patients with Multiple Sclerosis, *Stud Health Technol Inform* **257** (2019), 189-193.

[10] D.J. Buysse, C.F. Reynolds III, T.H. Monk, S.R. Berman, D.J. Kupfer, The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research, *Psychiatry Research* **28** (1989):193-213.

Role-Based Architecture for Secure Management of Telepathology Sessions

Rui JESUS^a, Pedro NUNES^a, Rui LEBRE^{a,b,1} and Carlos COSTA^a

^a *University of Aveiro, Portugal*

^b *University of A Coruña, Spain*

Abstract. Digital pathology is the computer technology that allows the management of the information generated by the whole-slide scanners from a microscopic slide, encompassing the virtual microscopy. This paper proposes and describes the architecture of a secure collaborative platform that integrates a web pathology viewer with role-based access control. The proposed architecture is ensured by a shared medical repository that serves web pathology viewer with the medical images, using the DICOM standard. The system offers collaborative work session management tools as the managing of users, sessions, access control to sessions, and many others. Furthermore, the use cases related to telepathology and e-learning are presented.

Keywords. WSI, collaborative, digital pathology, PACS, DICOM, access control

1. Introduction

Digital pathology has gained popularity in the most recent years as a consequence of the proliferation of the Whole-slide Imaging (WSI) [1,2], replacing the traditional microscopes [3]. The digital era in the pathology field brought new features in the review process: the samples can be easily accessed from anywhere and anytime and at the same time as other users and then remotely reviewed and annotated. The improvements lie in the diagnostic accuracy and efficiency of the support systems. The handling of the samples becomes more secure as there is no deterioration in the image quality [4].

Digital pathology made available new methods to teach histology and pathology that was impossible until the digital revolution [5]. The digital replaces the old and expensive microscopes and simplifies the process in a way that all that is required is a device with internet access and a web browser for concurrent access to the same sample, either for teachers or students. Still, the regulation nowadays forces to protect the patient's data [6]. Therefore, the handling of these samples over the network requires strong access control and security protocols.

Digital pathology made available new methods to teach histology and pathology that was impossible until the digital revolution [5]. The digital replaces the old and expensive microscopes and simplifies the process in a way that all that is required is a device with internet access and a web browser for concurrent access to the same sample, either for

¹ Corresponding Author, Campus Universitário de Santiago, Aveiro, Portugal; E-mail: ruilebre@ua.pt.

teachers or students. Still, the regulation nowadays forces to protect the patient's data [6]. Therefore, the handling of these samples over the network requires strong access control and security protocols.

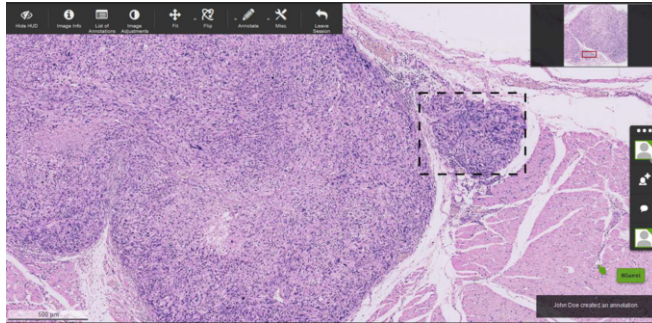


Figure 1. Screenshot of the platform during a work session.

This paper proposes and implements a secure architecture for the management of digital pathology remote working sessions². The management platform provides means for a session creator to restrict and personalize permissions for the access and handling of medical images. The framework also integrates a pure web pathology viewer fully compliant with Digital Imaging and Communications and Medicine (DICOM) standard, the Dicoogle open-source project to support the WSI storage, and a web platform for the management of all the working sessions and their associated users. Furthermore, the introduction of a security layer [7] enables to control access policies in a non-secure environment, deploying an innovator multi-repository concept. The data management is assured by this accounting mechanism specially designed for medical imaging archives which creates the concept of virtual archives for each user.

2. Background

2.1. DICOM-PACS Universe

The evolution of the radiology regarding the digital paradigm let to the implementation of Picture Archiving and Communication Systems (PACS). PACS is an aggregation of hardware and software services that orchestrate acquisition, storage, and display of medical imaging data. All of the previous steps are integrated into digital networks supported by DICOM standard. PACS rely on DICOM, which is one of the most popular standards in the medical imaging field [8].

In 2010, microscopy was introduced in the DICOM, as a result of the rapid spread and adoption of the standard. The publication of the supplement 145 boosted the development of the automated whole-slide scanners [4]. However, the technology maturity level and initial vendor-lock positioning of scanners industry does not promote a fast adoption of DICOM-WSI in the first years.

Whole-Slide Images are specimen-centric, unlike the traditional DICOM information model, that are considered patient-centric where the specimen is not the

² Demo: <http://demo.dicoogle.com/pathobox>; Demo Video: <https://youtu.be/Mmsb25edcOo>

most relevant subject [9]. The most recent DICOM standard supplement dedicated to WSI specify the workflows regarding the preparation, acquisition, and storage of this kind of imaging objects [9].

2.2. Dicoogle

Dicoogle³ is an open-source PACS [10]. Its applicability encompasses different contexts as clinical, research and academic. Dicoogle's architecture follows a modular approach, allowing the development of new features over time.

The Dicoogle Software Development Kit (SDK) allows the addition of new features by third parties without the need to change the software core. The Dicoogle SDK provides interfaces that the developers need to implement to build a plugin⁴. Dicoogle, during the startup, automatically loads the modules contained in this directory and all the operations related to storage, querying and indexation become immediately available via its API.

3. Architecture Overview

Figure 2 shows the general architecture of the proposed system. Four main components can be identified: the PACS archive, the management platform (MP), the web viewer and the access control module.

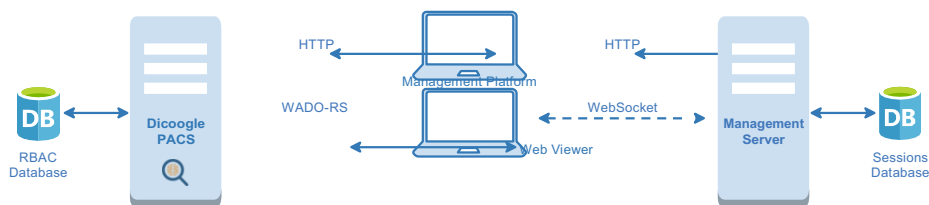


Figure 2. Collaborative platform general architecture.

In this context, a working session is a live persistent document, that maintains synchronization between all users. The MP is where all the data, regarding sessions, users and their permissions, is stored. This component is independent of the viewer, so with little effort, other viewers can be attached to the management platform to support the working sessions.

The purpose of the MP is to give the user the necessary tools to manage his sessions, with the possibility to create user groups (that speed up the process of inviting users), define user permissions in the working sessions, define access control configurations to each session, insert new case studies and create new sessions. The MP handles the access to the working sessions at two different levels: controlling the access to the working sessions through user-defined parameters, such as access dates, invited users and their permissions; controlling the access to the case studies that the sessions are based on.

The users can be invited via two ways: email, where a personal invitation link to the session is delivered, with preemptively defined permissions; or through a public share-

³ Website: www.digoogle.com; GitHub: <https://github.com/bioinformatics-ua/dicoogle>

⁴ <https://bioinformatics-ua.github.io/dicoogle-learning-pack>

able link, up to a defined limit, which will be given no permissions to edit the session. These permissions can be changed live in the session.

The permissions on the session define which actions the user can perform concerning the image and the study, such as moving around the image or create annotations, and regarding the collaborative tools, such as inviting or kicking users from the session.

The MP controls the users in the session, registering who and how many users are connected, preventing the simultaneous usage of the same invitation link and limiting the number of users connected through a public link.

Given that the platform works over medical repositories, including personal archives that constitute sensitive information and therefore need to be compliant with the General Data Protection Regulation (GDPR), the access to the case studies needs to be controlled. In this case, access to the case studies is role-based and controlled by an existing accounting mechanism developed by Lebre et al. [7]. This mechanism is role-based, preventing users from accessing archives they should not have access to. The permissions can be extended to third-party entities allowing to create multiple virtual archives and share them amongst other users and institutions.

Upon logging in into the MP, a role is given to the user, defining which case studies he/she has access to. The user can upload to the platform his cases or the system administrator can grant this user access to existing archives.

Within a working session, the viewer, using the web link, determines the visualization mode to be used, retrieving the user's permissions and session's details from the MP and the specific medical case from the PACS archive. It has three visualization modes, all sharing the same basic visualization tools: live mode, the representation of the working session; replay mode, allowing to review all the actions on the working session; archive mode, the study case can be viewed without any of the collaborative features.

4. Use Cases

A collaborative paradigm in pathology enables the interaction and discussion by several users simultaneously, each in their own device. The collaborative system was designed to satisfy two usage scenarios, clinical and educational.

In the clinical field, the leader of the study can create a group of users from his peers to an easier attribution of permissions. Alternatively, the leader may specify or change the permissions of each member when establishing the working session. The replay mode will allow the peers to review actions performed and be able to keep up with the research. In the educational field, the groups could be formed with students or can be given the public link, allowing these users to follow along with the class. By changing the permissions of one of the attendees, the educator can, for instance, evaluate the ability of a specific student to detect critical regions in the study. Moreover, tasks for students can be defined. For instance, annotation of a working case that will be saved as a session than can be reviewed later by the evaluator.

On both cases, the framework virtualizes the concept of PACS archive. I.e., the management platform creates virtual sessions where the users only have access to the DI-COM objects that the creator of the session has previously defined. However, all of the DICOM objects are stored in the same archive. The framework provides then, the

capacity to share studies and instances among users belonging to distinct healthcare organizations or departments.

5. Conclusion

In the medical field, the imaging field represents a crucial tool in the quality of diagnosis. Sharing patient data between entities boosts the accuracy of decision making. In this line, collaborative work is a breaking-change in clinical decision-making. The collaborative paradigm addresses also educational purposes, allowing collaboration among professors and students.

This paper presents the architecture and use cases of a DICOM web digital pathology viewer with access control mechanisms integrated. The proposed framework creates the concept of collaborative sessions. However, the existence of such a framework requires a permission management system which led to the deployment of a management framework that promotes the resource sharing among users or institutions. The working sessions are virtual web viewers where an administrator can define which users can have access to which resources. These features turn the system institution-free and enhance telemedicine and remote diagnosis. Moreover, using pure web technologies, the system can be used anywhere, anytime, regardless of the operating system and without any prior installation.

Acknowledgments

This work was funded by the FCT, Fundac o para a Cie ncia e a Tecnologia, under the project UID/CEC/00127/2019 and POCI-01-0145-FEDER-016385 (NETDIAMOND). This work has received support also from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Unions Horizon 2020 research and innovation programme and EFPIA.

References

- [1] L. Pantanowitz, P. Valenstein, A. Evans, K. Kaplan, J. Pfeifer, D. Wilbur & T. Colgan, Review of the current state of whole slide imaging in pathology, *Journal of Pathology Informatics* **2** (2011), 36
- [2] A. Saco, J. Bombi, A. Garcia, J. Ramrez & J. Ordi, Current status of whole-slide imaging in education. *Pathobiology* **83** (2016), 79-88.
- [3] M. Triola & W. Holloway, Enhanced virtual microscopy for collaborative education, *BMC Medical Education* **11** (2011), 4.
- [4] G. Bueno, M. Fern andez-Carrobles, O. Deniz & M. Garc a-Rojo, New trends of emerging technologies in digital pathology. *Pathobiology* **83** (2016), 61-69.
- [5] L. Pantanowitz, J. Szymas, Y. Yagi & D. Wilbur, Whole slide imaging for educational purposes, *Journal of pathology informatics* **3** (2012), 46.
- [6] K. Abouelmehdi, A. Beni-Hessane & H. Khaloufi, Big healthcare data: preserving security and privacy, *Journal of Big Data* **5** (2018), 1.
- [7] R. Lebre, L. Bastia o & C. Costa, Shared Medical Imaging Repositories, *Studies in Health Technology and Informatics* **247** (2018), 411-415.
- [8] J. Silva, T. Godinho, D. Silva & C. Costa, A community-driven validation service for standard medical imaging objects, *Computer Standards & Interfaces* **61** (2019), 121-128.
- [9] C. Daniel, F. Macary, M. Rojo, J. Klossa, A. Laurinavius, B. Beckwith & V. Della Mea, Recent advances in standards for Collaborative Digital Anatomic Pathology, *Diagnostic Pathology* **6** (2011), S1-S17).
- [10] F. Valente, L. Silva, T. Godinho & C. Costa, Anatomy of an Extensible Open Source PACS, *Journal of Digital Imaging* **29** (2016), 284-296.

Safer Surgery Checklist: Barriers in the Adoption of Tablets in Operating Room

Giuliana COLUSSI^{1a}, Gabriela GARCIA ^a, Maria GRANDE^a, Daniel LUNA^a

^a Health Informatics Department, Hospital Italiano de Buenos Aires, Argentina

Abstract. It is crucial that registering the checklist needs to be done synchronously and according to the care tasks within the operating room. Therefore, we aim to find out what are the difficulties that the surgical technicians have when checking the list and exploring potential barriers to implementing the use of tablets. Qualitative research was conducted based on focus groups, in-depth interviews and on-site observations until a full coverage of the topics was achieved. We detected that this problem includes technical aspects (connectivity, hardware and software issues), those related to workflow and logistics, those related with registration process, and those related to organization culture and power structure. Based on results, we developed a plan to solve problems detected.

Keywords. Checklist; Operative Surgical Procedure; Patient Safety; Qualitative Research

1. Introduction

In order to standardize care processes in the operating room, the World Health Organization designed a checklist with the advice of surgeons, anesthetists, nurses, and patient-safety experts from around the world [1]. It aims to reinforce established safety practices and promote communication and teamwork across multiple clinical disciplines.

Checklist's implementation during the surgical procedure is intended to work as a tool to improve patient safety, human factors, reduce the incidence of adverse events, morbidity and mortality [2] [3] [4].

However, evidence shows that there are some factors that prevent this procedure from being done on time and correctly [5], despite being a requirement of the Joint Commission International [6]. The barriers to the correct implementation of the surgical checklist are multifactorial and cover both cultural and structural aspects, including communication gaps, the benefit perceived by the user, lack of understanding of the procedures, the perceived ambiguity of the process and negative perceptions regarding operational efficiency, and others [7] [8].

In 2014 the Italian Hospital of Buenos Aires (HIBA), computerized the checklist to be used during the surgical process, and in 2015 the use of tablets was implemented so that the surgical technicians team can complete it from a mobile device, and thus be able to do it next to the patient. However, the use of the tablets was not maintained

¹ Corresponding author: Giuliana.colussi@hospitalitaliano.org.ar.

over time, and gradually ceased to be used for the checklist registration.

It is crucial that registering the checklist needs to be done synchronously and according to the care tasks within the operating room. Therefore, we aim to find out what are the difficulties that the surgical technicians have when checking the list and exploring potential barriers to implementing the use of tablets, in order to formulate solutions that guarantee the correct realization of it.

2. Methods

The HIBA is a center of high complexity located in the City of Buenos Aires. Every year performs 52,000 surgical procedures in its 41 operating rooms.

Qualitative research was conducted based on focus groups, in-depth interviews and on-site observations until a full coverage of the topics was achieved. Both the interviews and the focus groups were audio recorded for further analysis, protecting the identity of the participants who voluntarily participated through an informed oral consent. The research project was approved by the institutional ethics committee (CEPI # 5001). The study was performed in full agreement with current national and international ethical regulations.

Between November and December of 2018, 2 focus groups were held, one with the morning shift group and the other one with the afternoon shift group. A total of 41 surgical technicians participated. A tool from the Lean Change Management was used as a triggering material for the investigation. This tool consists in expressing which is the summoning problematic [9]. Then each member, individually, writes in a piece of paper the reasons why they think this is happening. Next step, the reasons are collected, grouped by categories and the different issues that arise begin to be discussed among all the participants. This way, participant's feedback is shared, eliminating biases and generating perceptions of greater value, where possible solutions arise from the group itself.

A group interview was carried out with 2 key participants: the surgical technicians coordinator and the management and update coordinator of the surgery department. Three surgical processes were observed on different days, schedules and operating rooms. Also electronic devices, access to networks and the software involved during the surgical processes were reviewed. The observations were carried out in the central operating room, covering all the tasks performed by surgical technicians, especially focusing on the checklist registration process. The information was systematized in field newspapers, elaborating a dense description [10] through interviews and focus groups.

The analysis process was carried out simultaneously with the information collection process, through the coding and categorization of information, based on a constant comparison process. The categories were built taking into account the central research questions such as the findings that emerged from the same focus groups, interviews and observations, building a descriptive framework that helps contextualize the collected data [11].

3. Results

Emerging categories according to the checklist completion arise from the combination of three employed techniques, that allowed greater wealth in the understanding of the phenomenon. We found three main dimensions. These dimensions account for technical aspects expressed in problems ranging from device's hardware, software, and problems with Wifi signal. We also found workflow related problems, manifested as logistic and organization issues, as well as problems related to the process at registering the checklist, to the organizational culture and about power structure.

3.1 Technical aspects

We evidenced difficulties or barriers in the inquiry that were related to technical aspects associated with checklist registration, which includes connectivity, software and hardware problems.

Connectivity: We detected connectivity problems as unstable range of Wifi inside each operating room, that causes logged out, browsing slow, or lose information loaded. This could make the registration process problematic and longer times.

"often the wifi is wrong and that's it, if you don't have Wi-Fi ... it doesn't work"

Software problems: In order to access checklist interface, surgical technicians must enter to the Electronic Health Record (EHR) from the tablet. In this sense the screen does not adapt to the resolution of used device, making zooming and scrolling necessary. This creates which creates possibilities of error when accidentally touching buttons and when they need to click many times to enter the surgical episode of each patient.

"It is cumbersome to have to enlarge and shrink the image to check option"

"I had to be enlarging the screen, you went to the next step and you had to enlarge it...if there was an application in a standard size, that's it, the size would not be a problem"

Hardware problems: Tablets were slow and usually freeze, and users have difficulty writing on them:

"The tablet's size creates difficulties for writing" "The icons looked very small"

3.2 Workflow

Within this dimension we find three barriers, one is related to dynamics inside the operating room, another refers to operational logistics of tablets and the last one refers to registration itself. Dynamics inside the operating room: in the record of the observations, we noted that operating rooms have 2 desktop computers, one is exclusively used by anesthesia team, expressed explicitly, and the other one, although it was not expressly announced, was constantly used by surgeons. This makes difficult to synchronize the checklist because surgical technicians must wait until computer was available.

*"getting to the computer is very difficult and the computer is far from patient",
"in order be able to register something that is done at a certain moment you need to get away
from the patient, turn their backs at them. It is
very careless and when they register the data, the computer is busy"*

Tablets operational logistics: responsibility generated by using a device that is not their

own but belongs to the institution is explained. In order to use them, a process to ensure safety and cleanliness of tablets had to be respected. Circuit becomes disruptive in routine surgical technicians tasks, since they have to go to storage and complete a form when borrowing the devices, taking responsibility for them, which consumed too much time and made them uncomfortable to continue with another tasks. In addition, tablets were not always available because they were stored without enough battery, and chargers were missing.

“you were going to ask for tablet, you opened it, it had no battery, you had to go back to ask for a charger, and there were no chargers”, “They (surgeons) asked for charger to another use (phone charge), I don’t know if they would take it home, or leave it there. Doctors have their radio, they never have battery and they have the same charger plug as tablets”, “A lot of liability for possible losses between surgeries while we have to prepare material”

Check List Registration: problematic of synchronous register of checklist is given by a set of topics related to assistencialist aspect of surgical technicians work. On one hand, presence of a mobile device such as tablets could hinder dynamics their work, since they need to have both hands free to specify their care tasks with patient, since interrogation until ensuring availability of necessary supplies on operating table.

“At patient’s admission we have many things to do, we cannot have it (tablets) in our hands”, “One difficulty is having to have it during surgeries”

On the other hand, check list is generic and unique for different services and types of surgeries, which implies that all fields present in it do not always apply, or some are missing, or order items may not correspond to a specific workflow.

3.3 Organizational culture and power structures in operating room

Transversally to addressed problems, an underlying category refers to organization culture and power structures in operating room. We observed that power structures within operating room set out priorities on devices use of such as computers and tablet chargers. Sometimes surgical technicians relegate tasks to comply with requests from surgeons in the ward, for example bringing their cell phone, taking a picture of surgical marking on patient. Likewise, although correct application of check list is a responsibility of the entire operating room team, surgical technicians adopt this task as their own.

4. Discussion and Conclusion

We explored difficulties of surgical technicians that arise checklist procedure. We detected that this problem includes technical aspects (connectivity, hardware and software issues), those related to workflow and logistics, those related with registration process, and those related to organization culture and power structure.

Based on preliminary results, we solved connectivity issues, we installed more Wi-Fi access points in operating rooms, and we developed a plan to replace obsolete devices. Regarding the software, we plan to develop a new mobile application that responds to specific registration needs of surgical technicians team. However, solving technical aspects would not solve all difficulties detected. Despite our plan with informatics issues, additionally, it will be convenient to attend organizational and

cultural factors, concordant with literature [8].

Workflow problems could be solved by installing a bedside post next to patient to hold tablets and allowed hands free. This would ensure that surgical technicians have a device exclusively for them, and able to prevent wasting time in withdrawal and return.

Finally, we consider that could be convenient to redesign checklist according to specific context in which it will be used, as Raman recommends [12]. According to organization culture and power structure, it would be desirable to work with the entire Department of Surgery to raise awareness among all members of operating room about the importance of correct completion of checklist on patient safety, following the importance of the team's commitment is highlighted [13], as a collective responsibility.

References

- [1] Manual de aplicación de la lista OMS de verificación de la seguridad de la cirugía 2009: la cirugía segura salva vidas [Internet]. World Health Organization; 2009. Available: https://apps.who.int/iris/bitstream/handle/10665/44233/9789243598598_spa.pdf;jsessionid=2649477FC0853868EFF22C770630D76E?sequence=1
- [2] Thomassen Ø, Storesund A, Sjøteland E, Brattebø G. The effects of safety checklists in medicine: a systematic review. *Acta Anaesthesiol Scand*. 2013;58: 5–18.
- [3] Hohenfellner R. Re: A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population [Internet]. *European Urology*. 2009. p. 395. doi:10.1016/j.eururo.2009.05.038
- [4] Weiser TG, Haynes AB, Dziekan G, Berry WR, Lipsitz SR, Gawande AA. Effect of A 19-Item Surgical Safety Checklist During Urgent Operations in A Global Patient Population [Internet]. *Annals of Surgery*. 2010. pp. 976–980. doi:10.1097/sla.0b013e3181d970e3
- [5] Tostes MF do P, Galvão CM. Surgical safety checklist: benefits, facilitators, and barriers in the nurses' perspective. *Rev Gaucha Enferm*. 2019;40: e20180180.
- [6] Uppot RN, McCarthy CJ, Haynes AB, Burk KS, Mills T, Trifanov DS, et al. A Verbal Electronic Checklist for Timeouts Linked to the Electronic Health Record. *J Am Coll Radiol*. 2017;14: 1322–1325.
- [7] Sparks EA, Wehbe-Janek H, Johnson RL, Smythe WR, Papaconstantinou HT. Surgical Safety Checklist compliance: a job done poorly! *J Am Coll Surg*. 2013;217: 867–73.e1–3.
- [8] Fourcade A, Blache J-L, Grenier C, Bourgain J-L, Minvielle E. Barriers to staff adoption of a surgical safety checklist. *BMJ Qual Saf*. 2012;21: 191–197.
- [9] Project Management Institute. *Managing Change in Organizations: A Practice Guide*. Project Management Institute; 2013.
- [10] Geertz C. *La interpretación de las culturas*. 1991.
- [11] Ellis C, Strauss A, Corbin J. Basics of Qualitative Research: Grounded Theory Procedures and Techniques [Internet]. *Contemporary Sociology*. 1992. p. 138. doi:10.2307/2074814
- [12] Raman J, Leveson N, Samost AL, Dobrilovic N, Oldham M, Dekker S, et al. When a checklist is not enough: How to improve them and what else is needed [Internet]. *The Journal of Thoracic and Cardiovascular Surgery*. 2016. pp. 585–592. doi:10.1016/j.jtcvs.2016.01.022
- [13] Singer SJ, Molina G, Li Z, Jiang W, Nurudeen S, Kite JG, et al. Relationship Between Operating Room Teamwork, Contextual Factors, and Safety Checklist Performance. *J Am Coll Surg*. 2016;223: 568–580.e2.

Sensor-Based Decision Support for the Allocation of Patient Attendants in Hospitals

Sascha GFELLER^a, Philip KYBURZ^a, Kai-Uwe SCHMITT^{b,c}, Thomas BÜRKLE^a,
Jürgen HOLM^a and Murat SARIYAR^{a,1}

^a*Bern University of Appl. Sciences, Dept. Medical Informatics, Switzerland*

^b*Bern University of Appl. Sciences, Dept. Health, Academic-Practice-Partnership, Switzerland*

^c*Insel Gruppe, University Hospital Bern, Switzerland*

Abstract. In hospitals, patient attendants are often necessary in order to closely monitor patients with high risk of self-endangering actions and reactions. However, such additional monitoring of patients is associated with high costs. In this paper, we describe a technical infrastructure for monitoring the patient's activities, which helps to assess whether an attendant should be requested. It was central to for us to use non-invasive sensors and to exploit a variety of patient data such as heart rate, micro-activity and oxygen saturation.

Keywords. clinical decision support; sensors; micro-activity; patient sitters

1. Introduction

Monitoring inpatients in hospitals is an ever-increasing challenge, as the amount of multimorbidity rises with aging population. Especially in acute care settings, many adverse events, such as delirium or a psychotic episode, are not related to the actual reason for being hospitalized. Patients who develop a delirium are disoriented, confused and/or unable to understand their current situation. In the elderly patients, a delirium can be superimposed on dementia. As a result, such patients might suddenly remove medical devices (such as venous access devices, tubes, catheters, electrodes, cables, etc.), they show stress reactions and/or attempt to stand up. Currently, in nursing practice, it is often difficult to predict whether and when a patient will develop such states that require close monitoring [1, 2]. There are certain known risk factors (such as age, drugs used, etc.), but often health professionals can only rely on their experience for taking any precautions or initiating any measures [3].

Attendants can be nursing trainees, medical students and non-health care professionals [4, 5]. Hospitals usually have a pool of such staff. Their job is to permanently watch the patient at risk and to intervene if required. On a normal ward, the options for interventions are limited to, e.g., calming down the patient by talking or calling additional nursing staff for help. Such additional monitoring of patients is

¹ Corresponding Author, Murat Sariyar, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: murat.sariyar@bfh.ch.

associated with high costs. According to internal information from a Swiss university hospital (approximate bed size 900) yearly costs for patient attendants can sum up to 60 full-time nurse equivalents.

As the request for an attendant must be made several hours in advance, a positive decision can result in attendants sitting beside a sleeping patient, thus wasting resources. A case study at a hospital in Ohio reported that using a risk-based digital assessment tool for regularly allocating attendants can reduce the working hours without increasing the adverse events, indicating that there is potential in reducing costs [6].

Our goal was to provide and examine a sensor-based technical infrastructure monitoring the patient's activities, which might be helpful to assess whether an attendant should be requested. It was central for us to use non-invasive sensors and to exploit a variety of patient data such as heart rate, micro-activity and oxygen saturation.

2. Methods

2.1 Review of the literature

We conducted a literature research in Cinahl, PubMed, Web of Science and Google Scholar with a focus on the work of patient attendants, the decision process of employing them, sensors for detecting and preventing adverse events, and clinical decision support systems for allocating patient attendants. Keywords were identified in particular through preliminary discussions with stakeholders. These were then entered in various combinations:

- "sitter" or "attendants"
- ("sitters" OR "attendants") AND "risks"
- "constant observers"
- "constant observers" AND "risks"
- "inpatient" AND "fall detection"
- "patient monitoring" AND "fall detection"
- "mattress sensor care"
- "health data sensor"
- "patient monitoring sensor"
- "vital data sensor"

2.2 Approaches considered

In order to get insights into the actual work of patient attendants, a semi-structured questionnaire for interviews with attendants was prepared and eight interviews were conducted. The questions were related to the preparation and training for the job, contents of the tasks tackled, subjective assessments of the job, and opinions to the use of sensors in the context of the job. Inclusion criteria for the interview were: working at least half a year as patient attendant within the last two years, speaking German or English, working in a Swiss hospital, being of age.

To derive the requirements of a sensor-based monitoring system we developed several use-cases. In addition, we composed scribbles inspired from the Design-Thinking approach to visualize the processes itself and potential problems from the perspective of a patient attendant [7]. Based on that input, we implemented a web

application to integrate the sensor data using the agile Scrum approach [8] in ten one-week sprints. We applied the JavaScript framework vue.js and loopback for the frontend and the backend, respectively.

3. Results

The results of our literature research showed a variety of definitions of the concept “patient attendant” [9,10]. Many different settings have been reported that require different forms of monitoring and interventions. This was corroborated by the eight interviews we have conducted. Some attendants just observed, others had to intervene, others used a lot of communication with the patient. In most cases, there were no trainings beforehand. Attendants had to deal with different medical conditions such as delirium, dementia, schizophrenia, suicidality, alcohol problems, etc. Interventions comprised calming down the patient, preventing the patient from standing up or make sure that a member of clinical staff is available swiftly. Approximately half of the deployments were futile because no interventions were required. The study participants indicated that sensors monitoring patient mobility and/or attempts to leave the bed might be beneficial.

We developed an evaluation matrix for sensors currently available on the market. Main criteria were measured parameters, running time, interfaces, how data is gathered (extend of invasiveness), costs, and certification as a medical device. Based on these attributes, we decided to use two complementary and nearly non-invasive sensors. One sensor is the *Mobility Monitor*™ (momo) of Compliant Concept, which monitors micro-movements of the patients in their bed by placing a sensor under the mattress. The system can detect accurately agitation and bed-leaving events. The other sensor (*Everion*™) from Biovotion is placed on the patients arm and measures pulse, skin temperature, oxygen saturation, respiratory rate, and some more parameters. For the momo sensor a REST API is available, for Everion a JAVA SDK has to be used in order to read the data.

The overall architecture of our prototype, called *SensorLink*, is given in Figure 1. Both sensors are queried every 10 seconds using the corresponding API. We combined the data of both sensors with clinical parameters into one single backend and provide interactive graphical outputs in order to assess whether a patient needs more monitoring. An example output of the front end is given in Figure 2. Interactivity in the graphical outputs is achieved by using the ApexCharts.js library for visualiations.

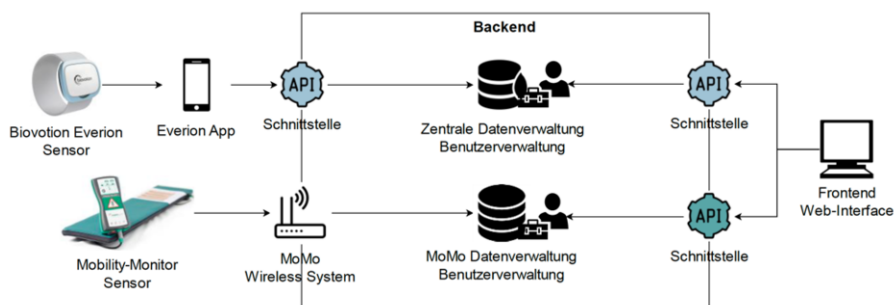


Figure 1. Overview on the prototypical sensor network system with two sensors (Everion and the Momo sensor) that provide data via different interfaces to the backend for display on a web interface.

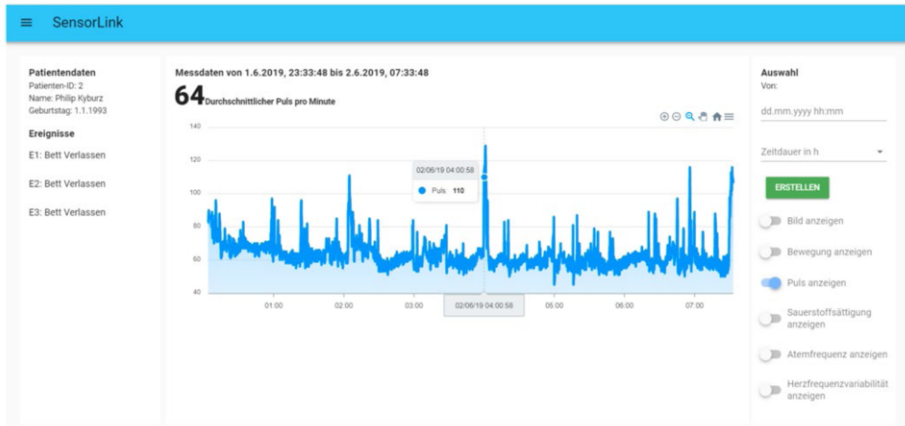


Figure 2. The graphical user interface of the SensorLink system. Here, the pulse trajectory of one night is shown in the main panel. The peak pulse value is at 110 and the average value throughout the night was 64.

The data in Figure 2 was collected by installing the sensors and the system at the medical informatics laboratory in Bienne/Switzerland with the aim to validate our solution. The co-authors SG and PK slept two days each in the laboratory. Central events, such as leaving the bed, are given at the left panel of the graphical interface. In the right panel, different criteria can be selected, e.g., measurement period, micro-activity, pulse, heart frequency variability, etc. The results for this selection are provided in the main panel. The graphics are interactive and aligned with each other based on the time stamps, which allows to monitor several parameters at desired time points.

4. Discussion

Currently, various sensor-based wearables, external sensors and even mobile video monitoring are available [11,12]. However, such solutions are frequently not apt for either reducing the amount of complications or for improving the decision efficiency regarding patient attendants' placements [13]. Especially, inappropriate decision and bad user experience associated with such additional systems pose serious problems [14]. To tackle these problems, non-invasive sensors data should be used together with rule-based reasoning and machine learning [15]. We have provided a first step in that direction.

The experiences in our project showed that it is rather cumbersome to individually integrate the diverse sensors. A decision-making system that uses many different sensors should use a comprehensive gateway solution that enables plug and play of those sensors. One commercial solution (<https://www.leitwert.ch>) provides an IoT middleware for managing and configuring such a gateway. It allows to collect data on a local server, which paves the way to use wearable medical devices in health care settings. Clocking of the different sensors is also a relevant issue, as the clock pulse can

vary between sensors. Further, it should be taken account that the amount of data generated by many sensors in short time intervals indicate that cloud-based solutions should be taken into consideration for scalability reasons.

We have gained first insights into the possible advantages of increased information on patients and suitable graphical summaries for the decision concerning placement of patient attendants. It is crucial, however, to prove the benefits of such a decision support tool for those who decide in practice. This requires a proper implementation study to evaluate the impact in a real-world setting. Central outcomes are acceptability, adoption, appropriateness, costs, feasibility, and sustainability [16]. Our goal is to conduct an implementation study with an extended version of our prototype, using a sensor gateway and machine learning components that do not just condense information but deliver advice whether a request of a patient attend is adequate or not.

References

- [1] K. Harper, A. Barton, G. Arendts et al., Failure of falls risk screening tools to predict outcome: a prospective cohort study, *Emergency Medicine Journal* **35** (2018), 28–32.
- [2] J. De Wand, Delirium Screening: A Systematic Review of Delirium Screening Tools in Hospitalized Patients, *The Gerontologist* **55** (2015), 1079–1099.
- [3] F. Carr, The role of sitters in delirium: an update, *Can Geriatr J* **16** (2013), 22–36.
- [4] H. Tzeng, C. Yin, J. Grunawalt, Effective assessment of use of sitters by nurses in inpatient care settings, *Journal of Advanced Nursing* **64** (2008), 176–83.
- [5] M. Feil, S.C. Wallace. The Use of Patient Sitters to Reduce Falls: Best Practices, *Pa Patient Saf Advis* **11** (2014), 8–14.
- [6] D. Long, J. Dennis, Reducing sitter use in acute medicine while maintaining safety and quality, *Journal of Nursing Education and Practice* **9** (2018), 61–67.
- [7] M. Altman, T. Huang, J. Breland, Design Thinking in Health Care, *Preventing Chronic Disease* **15** (2018), E117.
- [8] H. Cervone, Understanding agile project management methods using scrum, *OCLC Syst Serv* **27** (2011), 18–22.
- [9] A. Schoenfisch, L. Pompeii, H. Lipscomb et al., An urgent need to understand and address the safety and well-being of hospital ‘sitters’, *American Journal of Industrial Medicine* **58** (2015), 1278–1287.
- [10] S. Solimine, J. Takeshita, D. Goebert et al., Characteristics of Patients with Constant Observers, *Psychosomatics* **59** (2018), 67–74.
- [11] T. Shany, S. Redmond, M. Narayanan et al., Sensors-Based Wearable Systems for Monitoring of Human Movement and Falls, *IEEE Sensors Journal* **12** (2012), 658–670.
- [12] P. Burtson, L. Vento, Sitter Reduction Through Mobile Video Monitoring: A Nurse-Driven Sitter Protocol and Administrative Oversight, *The Journal of Nursing Administration* **45** (2015), 363–372.
- [13] N. Kosse, K. Brands, J. Bauer et al., Sensor technologies aiming at fall prevention in institutionalized old adults: A synthesis of current knowledge, *International Journal of Medical Informatics* **82** (2013), 743–52.
- [14] M. Harrison, R. Koppel, S. Bar-Lev, Unintended Consequences of Information Technologies in Health Care: An Interactive Sociotechnical Analysis, *J Am Med Inform Assoc* **14** (2007), 542–9.
- [15] T. Syeda-Mahmood, Role of Big Data and Machine Learning in Diagnostic Decision Support in Radiology, *J Am Coll Radiol* **15** (2018), 569–576.
- [16] H. Christie, S. Bartels, L. Boots et al., A systematic review on the implementation of eHealth interventions for informal caregivers of people with dementia, *Internet Interv* **13** (2018), 51–59.

Smart Ageing: Digital Solutions for Future Care

Anthony MAEDER^{a,1} and Gary MORGAN^a

^a*College of Nursing & Health Sciences, Flinders University, Adelaide, Australia*

Abstract. We propose a framework for discourse on digital solutions to support consumers and carers in delivery of health care and services for aged persons, based on a major needs analysis conducted across 56 diverse business entities in Australia. The resulting framework was based on two major identified domains: "Ageing in Place" for independent living situations, and "Ageing with Care" for managed aged care facilities. The paper describes the process used and the intermediate outcomes which enabled the framework to be synthesized. It is anticipated that the framework could be used to inform future scoping studies and to enable collaborative design, implementation and delivery of appropriate smart ageing digital solutions.

Keywords. Ageing, health services delivery, models of care

1. Introduction

It is well established that the age profile of world population is shifting rapidly towards a significantly higher proportion of aged persons than previously. The World Economic Forum has estimated that by 2050 more than 20% of world population will be aged 60 years and older [1]. In Australia (typical of developed countries) the proportion is even higher: it has been estimated that by 2046, 22% of the population will be aged 65 years and older, and that average life expectancy will be 85 years with 19% of older people (i.e. over 4% of population) being aged 85 years or older [2].

This change in the essential nature of our population will have profound impact, stimulating transformational changes underlying health care mechanisms. Today's older citizens wish to remain living in their own homes and to pursue an active lifestyle as long as possible, with appropriate support where necessary [3]. Those who require assisted living and clinical management through the health care system, expect that services reforms which will ensure optimal quality of life can be enjoyed in those circumstances [4]. We need to develop a variety of approaches to deal with this transformation, some of which will undoubtedly incorporate digital technologies.

Prior work has provided landscape commentary on opportunities for technology-delivered health care support for ageing citizens by appropriate technologies [5] and from clinical perspectives [6]. A common emphasis in these domains is the maintenance of independence and quality of life for subjects of care [7]. The consequential impact of these trends on health technology markets has also been characterized [8].

The research study summarised here was intended to provide a qualitatively validated clarification of major health services delivery needs for individuals of a future

¹ Corresponding author: A. Maeder, Email: anthony.maeder@flinders.edu.au

ageing society which can be potentially addressed by digital solutions. These major needs may be identified by the subjects-of-care themselves, or may be known to the providers of care services who deal directly with individuals: both groups constitute a body of end-users for the needs-derived solutions.

Before developing such solutions as these, it is important to understand their scope and to leverage common elements of them to derive the best outcomes. This could be expediently achieved by using a high level framework for considering approaches to developing digital solutions such as these: the major end-user needs identified here provide a basis for describing such a framework. We do not claim uniqueness for the framework but we believe that if similar exercises were undertaken, there would likely be a degree of commonality due to the diversity of the stakeholder sectors represented.

2. Methods

During February to October 2017 structured interviews were conducted with 56 Australian business entities representing a wide cross section of organisations involved with services for ageing. These included companies and not-for-profits engaged in community and aged care, hospital and acute care, primary care, nursing and allied health, finance, insurance, technology, software, built environment and urban design. Further consultations involved local and state government departments and peak professional bodies operating in the ageing sector, as well as ageing consumers directly.

Interviews were commenced with a short description of the context and purpose of the study. Interactive interviews followed, with a conversational approach adopted to enable free ranging inputs to be collected. Interview durations ranged from 30 mins to 2 hrs, and in most cases were conducted 1-to-1 between one of the authors and an informed senior executive representing the participating organisation. In some cases a second interviewee was present but in all cases the vast majority of responses were obtained from the primary participant.

A pro forma interviewing instrument based on prior work was developed to elicit problem statements and responses. The instrument contained the following discussion prompts:

- alignment of organisation with Ageing population sectors and situations;
- perceived major business problem statement and significance to organisation;
- actions currently being considered or already taken to address the problem;
- success criteria and need for collaboration and resources to progress solution;
- potential adoption and deployment settings within the organisation portfolio;
- strategic perspective on business impact of solving problem (2-5-10 years).

Conversational responses were captured verbatim from participants in written format. These were subsequently independently summarised using a simplifying terminology derived by a commonly used wordcloud formulation method, which extracted essential and repeated content from the conversational records. These condensed responses were subjected to topic clustering using concept-distance measures associated with the terminology, to extract common themes.

3. Results

From the topic clustering process, three distinctive end-user sectors were identified, which constitute domains of mutual interests spanning the overall ecosystem for a future ageing society:

Individual consumers and their personal circle of other persons, who are self-managing their ageing with support from both informal (e.g. family or friends) and professional (e.g. allied health workers) carers and the wider community within which they are situated (e.g. religious or social groups).

Community and aged care provider organisations and numerous associated service delivery agencies who are addressing and managing the needs of ageing clients, through services delivered in the home (e.g. cleaning, excercise), or through services delivered in residential managed care settings for those clients unable to care for themselves (e.g. feeding, medication).

Corporate or Government organisations, including clinical professionals, suppliers of health systems and technology solutions, developers and funders/payers responsible for business and infrastructure design and planning, and policy makers/influencers (e.g. bureaucracies, consultancies).

Two major thematic areas cutting across all three domains were discerned for development of digital technology solutions, from categorisation of the problem statements and significant recurring content descriptions from the interviews:

Ageing in Place, where consumers wish to continue living independently, in their own accommodation setting and maintaining responsibility for their own health status by accessing external sources of support. This will contribute to extending the years of healthy ageing and quality of life, as well as retaining longer engagement of ageing persons generally in economic and societal activities.

Ageing with Care, where it is necessary for consumers to live in conditions of clinical management, within specialised ageing facilities having associated inhouse health care professional support. This will provide platform systems for enhancement of key business processes and capabilities, improving consumer orientation of the current community and aged care business landscape, and aligning with needs to operate in a digitally informed knowledge environment.

For Ageing in Place, the dominant needs for solutions were identified as follows:

- Personal Wellness, with information gathered from about daily health related lifestyle habits of an individual, as well as summary data from digital monitoring sources and records of health check events.
- Community Connections, with a social networking online environment to enable individuals to establish and maintain a personalised set of community based group interactions and event participations.
- Adverse Event Reduction, using a “Health Smart Home” platform system for continuous surveillance of occupants, with multimodal sensing and pattern analysis to prompt alerts if anomalous situations are detected.

For Ageing with Care, the dominant needs for solutions were identified as follows:

- Personal Health Trajectories, representing and modelling of an individual’s ‘health trajectory’ based on health records and contributed data, mapping their advance through the ageing life course.

- Consumer Care Portal, a software system integrating information across service providers aligned with health data derived client profiles, and sharing health history and choices with clients/carers. Chronic Disease Management, via a generic software system including data collection and management, analysis of trends and recommendations on actions related to health condition self-care.

These findings are summarised in Table 1 below. While there could have been further less prominent topic areas extracted, to perhaps constitute a secondary layer of minor elements in the framework, the strength of consensus achieved by those identified here was sufficient to confirm them as dominant.

Table 1. Ageing digital technology solutions framework

Domain	Ageing in Place	Ageing with Care
Individual / Patient	Personal Wellness	Personal Health Trajectory
Community / Carers	Community Connections	Consumer Care Portal
Corporate / Clinicians	Adverse Event Reduction	Chronic Disease Management

4. Discussion

The focus of this work was an approach to enable us to identify and describe the needs and issues of older people and care providers, and which allows us to work collaboratively to design, implement and deliver solutions to those who need to apply them or use them. Through the coordination of research and development activities addressing these needs across diverse disciplines, organisations and sectors, innovative solutions can be developed which will create impact at scale.

This framework approach provides a set of channeled elements through which can be addressed the challenge of selecting and developing appropriate digital technology solutions to achieve the following high value benefits which participants indicated:

- Optimise the health and support the independence of older people, and thereby reduce the cost burden of care, by designing integrated solutions that focus on the person and their individual health circumstances.
- Empower older citizens to self-manage their health while ageing, to the greatest extent possible, with improved wellness and enablement while living independently and in communities, and provide for their eventual care in health facilities by equipping the clinical sector with integrated and comprehensive toolsets for appropriate and timely care.
- Address the business transformation needs of aged care and services providers, using living laboratory co-design models, to ensure solutions are appropriate, usable by customers and the workforce, and able to be integrated into wider systems, and validate their effectiveness at population scale such that they can be deployed widely.
- Develop transformational ‘ageing well’ approaches across the continuum of wellbeing, health and care services, which will improve the quality of life for ageing individuals and more generally enrich communities through the increased engagement of a previously excluded sector in the ageing population.

Feedback from interviewees included some broader aspirational elements as well as the specific needs analysis, which could not be categorized as precisely. These included adopting holistic viewpoints on health, living meaningful and purposeful lives, balancing increasing consumer expectations against diminishing services and choice, difficult to navigate the health system, complexity of integrating technologies.

5. Conclusions

The abovementioned benefits are consequential on the provision of appropriate technology solutions in the particular care setting. In the case of Ageing in Place, the wide range of potential technologies [9] must be tempered by their careful selection and application, to ensure success in their adoption and achievement of utility for purpose [10]. Further consideration will be necessary to address the challenges of data management and systems architecture [11] in this type of setting, as these are still far from being standardised. In the case of Ageing with Care, the overarching issue of integration and interoperability of component sub-systems and data flows [12] must be solved, harmonizing new solutions and current systems to enable care collaboration [13]. The Australian focus of this research may be seen as a limitation, but should nevertheless be comparable with the situations of many other high-income countries. Constructing a framework distilled from many lengthy and detailed interviews was seen as a beneficial way to provide scaffolding to support repetition of the process elsewhere. This approach should be of strong interest to the community and aged care sector to enable high level strategic planning and prompt social and economic flow-on impacts to its supporting sectors, to the health care system broadly, and finally to society at large.

References

- [1] *Global Population Ageing: Peril or Promise?*, <https://web.archive.org/web/20121122065922/http://www.icaa.cc/wef/Global-Population-Ageing-Book.pdf>, World Economic Forum, Global Agenda Council on Ageing Society, 2012
- [2] *Older Australia at a Glance*, <https://www.aihw.gov.au/reports/older-people/older-australia-at-a-glance/contents/summary>, Australian Government Department of Health and Ageing, Australian Institute of Health and Welfare, 2020.
- [3] *An Ageing Australia: Preparing for the Future*. <http://www.pc.gov.au/research/completed/ageing-australia/ageing-australia-overview.pdf>, Australian Government Productivity Commission, 2013.
- [4] *Caring for Older Australians*, <http://www.pc.gov.au/inquiries/completed/aged-care/report>, Australian Government Productivity Commission, 2011.
- [5] S. Koch, Healthy ageing supported by technology—a cross-disciplinary research challenge, *Informatics for Health and Social Care* **35** (2010); 81-91.
- [6] I. Vedel, S. Akhlaghpour, I. Vaghefi, H. Bergman, I. Lapointe. Health information technologies in geriatrics and gerontology: a mixed systematic review, *Journal of the American Medical Informatics Association* **20** (2013), 1109–19.
- [7] C. Siegel, T. Dorner, Information technologies for active and assisted living - influences to the quality of life of an ageing society, *International Journal of Medical Informatics* **100** (2017), 32-45.
- [8] L. Orlov, Technology for Aging in Place Market Overview, <https://www.ageinplacetech.com/page/market-overview-technology-older-adults> (2019).
- [9] M. Morris, B. Adair, K. Miller, E. Ozanne, R. Hansen, A. Pearce et al., Smart-home technologies to assist older people to live well at home, *Journal of Aging Science* **1** (2013), 1-9.
- [10] S. Fischer, D. David, B. Crotty, M. Dierk, C. Safran. Acceptance and use of health information technology by community-dwelling elders, *International Journal of Medical Informatics* **83** (2014), 624-35.
- [11] A. Maeder, P. Williams, Health smart homes: new challenges, *Studies in Health Technology and Informatics* **245** (2017), 166-9.
- [12] A. Georgiou, A. Marks, J. Braithwaite, J. Westbrook, Gaps, disconnections, and discontinuities - the role of information exchange in the delivery of quality long-term care, *Gerontologist* **53** (2013), 770–9.
- [13] S. Koch. Achieving holistic health for the individual through person-centered collaborative care supported by informatics. *Healthcare Informatics Research* **19** (2013), 3-8.

Statistically Prioritized and Contextualized Clinical Decision Support Systems, the Future of Adverse Drug Events Prevention?

Emmanuel CHAZARD^{a,1}, Jean-Baptiste BEUSCART^a, Michaël ROCHOY^b, Olivia DALLEUR^c, Bertrand DECAUDIN^d, Pascal ODOU^d, and Grégoire FICHEUR^a

^a Univ. Lille, CHU Lille, CERIMEA2694, F-59000 Lille, France

^b Univ. Lille, CERIMEA2694, General practitioner, F-59000 Lille, France

^c Pharmacy Department, Université catholique de Louvain, 1200, Brussels, Belgium.

^d Univ. Lille, CHU Lille, EA7365 GRITA, F-59000 Lille, France

Abstract. Clinical decision support systems (CDSS) fail to prevent adverse drug events (ADE), notably due to over-alerting and alert-fatigue. Many methods have been proposed in the literature to reduce over-alerting of CDSS: enhancing post-alert medical management, taking into account user-related context, patient-related context and temporal aspects, improving medical relevance of alerts, filtering or tiering alerts on the basis of their strength of evidence, their severity, their override rate, or the probability of outcome. This paper analyzes the different options, and proposes the setup of SPC-CDSS (statistically prioritized and contextualized CDSS). The principle is that, when a SPC-CDSS is implemented in a medical unit, it first reuses actual clinical data, and searches for traceable outcomes. Then, for each rule trying to prevent this outcome, the SPC-CDSS automatically estimates the conditional probability of outcome knowing that the conditions of the rule are met, by retrospective secondary use of data. The alert can be turned off below a chosen probability threshold. This probability computation can be performed in each medical unit, in order to take into account its sensitivity to context.

Keywords. Adverse drug events, Clinical decision support systems, data reuse.

1. Introduction

Adverse drug events (ADEs) are the most common type of iatrogenic injury. It is commonly admitted that coupling a clinical decision support system (CDSS) with a computerized order entry system (CPOE) may prevent ADEs [1]. But for current CDSS, over-alerting is an important issue: too numerous and inappropriate alerts may interrupt the clinicians' workflow and induce alert-fatigue [2,3], anger or annoyance [4]. This may prevent CDSS from improving patient safety [3,5]. As a consequence, up to 96% of alerts are overridden by prescribers [2,6–10], mainly because of poor appropriateness [11,12]. Unfortunately, appropriate alerts may be perceived as useless [2] and then overridden, but followed by actual ADEs [13–16]. The objective of this work is to propose a new approach to improve the ability of CDSS to prevent ADEs.

¹ Corresponding Author: emmanuel.chazard@univ-lille.fr

2. Material and methods

We first performed a narrative review of the tracks that have been proposed in the scientific literature, to reduce the over-alerting and alert-fatigue of CDSS, by searching for scientific papers in the Medline database, without time limit (keywords: CDSS, alert-fatigue, over-alerting). We also performed a recursive search including papers cited by those papers, and additional keywords. We then classified and discussed the tracks exposed in those papers. We finally proposed a new approach named SPC-CDSS (see below), in accordance with the results.

3. Results

3.1. Narrative review

In order to decrease over-alerting and alert-fatigue, it has been proposed to improve the way the CDSS interacts with the prescriber, and notably to improve the way the alerts are displayed [4], to provide the users with relevant instructions for post-alert medical management [3,17–19], and to involve non-medical healthcare professionals [6]. Some works also proposed to take into account the context of the drug prescription: user-related context (task, workflow, knowledge, preferences, medical unit) [2,20–22] and patient-related context (demographic data, risk factors) [3,17,20,21]. The temporal aspects have been identified as an important aspect: half of the prescribing errors occur on the first day of stay [14], the repetition of alerts has to be handled [9,20,21,23], as well as the kinetic of laboratory parameters [24].

It has also been proposed to improve the medical relevance of rules. One important track is to filter or tier the alerts by taking into account the strength of evidence [3,20,21]. It has been also proposed to ask medical experts to tier the alerts function of the severity of the potential outcome [3,5,6,20,21,25–27]: this approach seems to bring good results [5,25,27], but the experts hardly agree on the way to classify the rules [5,28]. Another option is to turn off rules that have the highest override rates [5,20,21,29]: this is efficient and mechanically decreases the override rate [29], but other studies demonstrated that most alert overrides were inappropriate [13–16].

Finally, given an ADE prevention rule, the conditional probability of ADE knowing that the conditions of the rule are met could be used to turn a rule on or off. This feature is largely asked by the physicians [3,20,21]. However, those approaches are still not implemented, although the reuse of electronic health records could enable to automatically estimate the conditional empirical probabilities of ADEs [30–33].

3.2. Proposal of SPC-CDSS

We propose to implement SPC-CDSS, which stands for “Statistically prioritized and contextualized clinical decision support systems”. The main idea is that, when a SPC-CDSS is implemented in a medical unit, it first reuses actual clinical data, and automatically searches for traceable outcomes (e.g. “INR>5”). Then, for each rule trying to prevent this outcome (e.g. “vitamin K antagonist & quinolone → risk of increased INR”), the SPC-CDSS automatically estimates the conditional probability of outcome, knowing that the conditions of the rule are met. In case of low probability

(the threshold can be customized), the corresponding alerts are automatically turned off (“statistically prioritized”). Moreover, this computation is performed in each medical unit, in order to take into account its sensitivity to context (“statistically contextualized”). This notion of context is a statistical proxy for latent undocumented variables, such as patients’ characteristics (e.g. some conditions, admission ground), organizational characteristics (e.g. monitoring policies), and physicians’ characteristics (e.g. knowledge about drugs, specialty, risk aversion) [30].

4. Discussion

Being able to compute the empirical probability of outcome for each ADE prevention rule may allow for two main benefits, that have been tested [30,33,34]. The first benefit could be to turn off some rules which empirical probability is below a chosen threshold (“A” arrow on Figure 1). Many CDSS rules are inappropriate [11,12], but experts hardly agree on which rule should be turned off [5,28]. The second benefit could be to show empirical evidence to the physicians to improve their adherence to the remaining alerts (“B” arrow on Figure 1). Indeed, too many appropriate rules are overridden by users [13–16]. Moreover, such information is requested by many physicians [3,20,21].

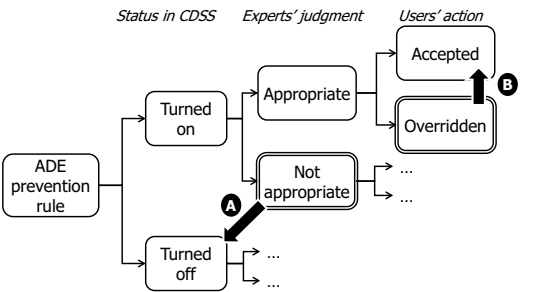


Figure 1. Expected benefits of SPC-CDSS: (A) turning some rules off, and (B) making rules more acceptable

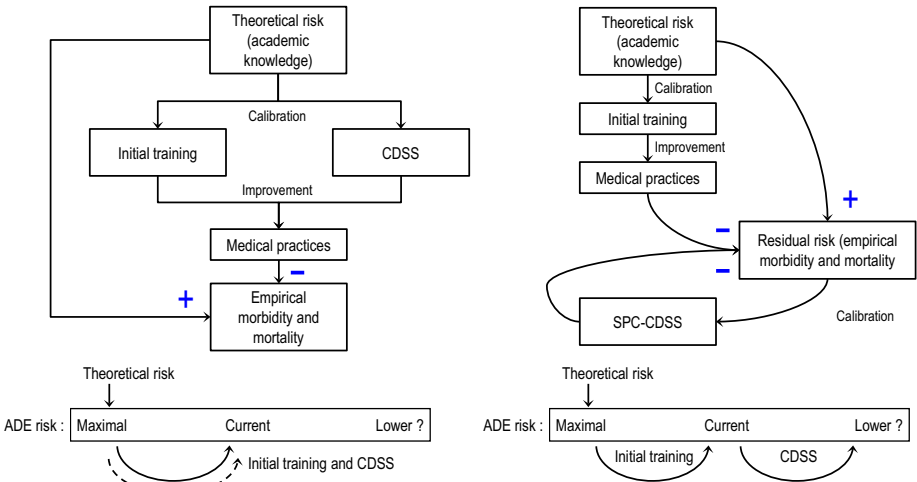


Figure 2. Left: current paradigm for ADE prevention. Right: proposed paradigm for ADE prevention.

The current paradigm for ADE prevention relies on the idea that CDSS should be calibrated based on the current academic knowledge (left part of [Figure 2](#)). But the physicians have also been initially trained on the basis of this knowledge. Consequently, the CDSS alerts are redundant with their own knowledge, and are useless. We propose to consider that the empirical morbidity and mortality are a residual risk, i.e. the maximal theoretical risk minus the benefits of the physicians' training. We then propose (right part of [Figure 2](#)) to use those empirical probabilities to calibrate the SPC-CDSS. Then, the SPC-CDSS alerts would be able to handle situations that have not been properly prevented by the initial training of the physicians. In terms of ADE risk, these considerations are summarized at the bottom of [Figure 2](#).

The present proposal should be tested and evaluated using actual clinical data. It also rises new issues, such as the threshold to choose, the variability between physicians, and the medical staff turnover. Finally, other solutions may be proposed to consider the final user's point of view.

References

- [1] A.X. Garg, N.K.J. Adhikari, H. McDonald, M.P. Rosas-Arellano, P.J. Devereaux, J. Beyene, J. Sam, and R.B. Haynes, Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review, *JAMA J. Am. Med. Assoc.* **293** (2005) 1223–1238. doi:10.1001/jama.293.10.1223.
- [2] H. van der Sijs, J. Aarts, A. Vulto, and M. Berg, Overriding of drug safety alerts in computerized physician order entry, *J. Am. Med. Inform. Assoc. JAMIA*. **13** (2006) 138–147. doi:10.1197/jamia.M1809.
- [3] S. Phansalkar, A. Desai, A. Choksi, E. Yoshida, J. Doole, M. Czochanski, A.D. Tucker, B. Middleton, D. Bell, and D.W. Bates, Criteria for assessing high-priority drug-drug interactions for clinical decision support in electronic health records, *BMC Med. Inform. Decis. Mak.* **13** (2013) 65. doi:10.1186/1472-6947-13-65.
- [4] M. Jung, A. Hoerbst, W.O. Hackl, F. Kirrane, D. Borbolla, M.W. Jaspers, M. Oertle, V. Koutkias, L. Ferret, P. Massari, K. Lawton, D. Riedmann, S. Darmoni, N. Maglaveras, C. Lovis, and E. Ammenwerth, Attitude of physicians towards automatic alerting in computerized physician order entry systems. A comparative international survey, *Methods Inf. Med.* **52** (2013) 99–108. doi:10.3414/ME12-02-0007.
- [5] H. van der Sijs, J. Aarts, T. van Gelder, M. Berg, and A. Vulto, Turning off frequently overridden drug alerts: limited opportunities for doing it safely, *J. Am. Med. Inform. Assoc. JAMIA*. **15** (2008) 439–448. doi:10.1197/jamia.M2311.
- [6] P.L. Smithburger, M.S. Buckley, S. Bejian, K. Burenheide, and S.L. Kane-Gill, A critical evaluation of clinical decision support for the detection of drug-drug interactions, *Expert Opin. Drug Saf.* **10** (2011) 871–882. doi:10.1517/14740338.2011.583916.
- [7] H. van der Sijs, A. Mulder, T. van Gelder, J. Aarts, M. Berg, and A. Vulto, Drug safety alert generation and overriding in a large Dutch university medical centre, *Pharmacoepidemiol. Drug Saf.* **18** (2009) 941–947. doi:10.1002/pds.1800.
- [8] C.-P. Lin, T.H. Payne, W.P. Nichol, P.J. Hoey, C.L. Anderson, and J.H. Gennari, Evaluating clinical decision support systems: monitoring CPOE order check override rates in the Department of Veterans Affairs' Computerized Patient Record System, *J. Am. Med. Inform. Assoc. JAMIA*. **15** (2008) 620–626. doi:10.1197/jamia.M2453.
- [9] T. Isaac, J.S. Weissman, R.B. Davis, M. Massagli, A. Cyrulik, D.Z. Sands, and S.N. Weingart, Overrides of medication alerts in ambulatory care, *Arch. Intern. Med.* **169** (2009) 305–311. doi:10.1001/archinternmed.2008.551.
- [10] M.-L. Yeh, Y.-J. Chang, P.-Y. Wang, Y.-C.J. Li, and C.-Y. Hsu, Physicians' responses to computerized drug-drug interaction alerts for outpatients, *Comput. Methods Programs Biomed.* **111** (2013) 17–25. doi:10.1016/j.cmpb.2013.02.006.
- [11] A.B. Taegtmeier, G.A. Kullak-Ublick, N. Widmer, V. Falk, and A. Jetter, Clinical usefulness of electronic drug-drug interaction checking in the care of cardiovascular surgery inpatients, *Cardiology*. **123** (2012) 219–222. doi:10.1159/000343272.
- [12] D. Fritz, A. Ceschi, I. Curkovic, M. Huber, M. Egbring, G.A. Kullak-Ublick, and S. Russmann, Comparative evaluation of three clinical decision support systems: prospective screening for medication errors in 100 medical inpatients, *Eur. J. Clin. Pharmacol.* **68** (2012) 1209–1219. doi:10.1007/s00228-012-1241-6.
- [13] S.P. Slight, D.L. Seger, K.C. Nanji, I. Cho, N. Maniam, P.C. Dykes, and D.W. Bates, Are we heeding the warning signs? Examining providers' overrides of computerized drug-drug interaction alerts in primary care, *PloS One*. **8** (2013) e85071. doi:10.1371/journal.pone.0085071.

- [14] T. Caruba, I. Colomбет, F. Gillaizeau, V. Bruni, V. Korb, P. Prognon, D. Bégué, P. Durieux, and B. Sabatier, Chronology of prescribing error during the hospital stay and prediction of pharmacist's alerts overriding: a prospective analysis, *BMC Health Serv. Res.* **10** (2010) 13. doi:10.1186/1472-6963-10-13.
- [15] K.C. Nanji, S.P. Slight, D.L. Seger, I. Cho, J.M. Fiskio, L.M. Redden, L.A. Volk, and D.W. Bates, Overrides of medication-related clinical decision support alerts in outpatients, *J. Am. Med. Inform. Assoc. JAMIA*. **21** (2014) 487–491. doi:10.1136/amiainjnl-2013-001813.
- [16] H. van der Sijs, T. van Gelder, A. Vulto, M. Berg, and J. Aarts, Understanding handling of drug safety alerts: a simulation study, *Int. J. Med. Inf.* **79** (2010) 361–369. doi:10.1016/j.ijmedinf.2010.01.008.
- [17] J.D. Duke, and D. Bolchini, A successful model and visual design for creating context-aware drug-drug interaction alerts, *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* **2011** (2011) 339–348.
- [18] A. Floor-Schreuderling, P.A.G.M. De Smet, H. Buurma, S. Amini, and M.L. Bouvy, Clarity and applicability of drug-drug interaction management guidelines: a systematic appraisal by general practitioners and community pharmacists in the Netherlands, *Drug Saf. Int. J. Med. Toxicol. Drug Exp.* **34** (2011) 683–690. doi:10.2165/11587270-000000000-00000.
- [19] J.D. Duke, X. Li, and P. Dexter, Adherence to drug-drug interaction alerts in high-risk patients: a trial of context-enhanced alerting, *J. Am. Med. Inform. Assoc. JAMIA*. **20** (2013) 494–498. doi:10.1136/amiainjnl-2012-001073.
- [20] E. Ammenwerth, W.O. Hackl, D. Riedmann, and M. Jung, Contextualization of automatic alerts during electronic prescription: researchers' and users' opinions on useful context factors, *Stud. Health Technol. Inform.* **169** (2011) 920–924.
- [21] D. Riedmann, M. Jung, W.O. Hackl, W. Stühlinger, H. van der Sijs, and E. Ammenwerth, Development of a context model to prioritize drug safety alerts in CPOE systems, *BMC Med. Inform. Decis. Mak.* **11** (2011) 35. doi:10.1186/1472-6947-11-35.
- [22] R. Marcilly, N. Leroy, M. Luyckx, S. Pelayo, C. Riccioli, and M.-C. Beuscart-Zéphir, Medication related computerized decision support system (CDSS): make it a clinicians' partner!, *Stud. Health Technol. Inform.* **166** (2011) 84–94.
- [23] G.J. Kuperman, A. Bobb, T.H. Payne, A.J. Avery, T.K. Gandhi, G. Burns, D.C. Classen, and D.W. Bates, Medication-related Clinical Decision Support in Computerized Provider Order Entry Systems: A Review, *J. Am. Med. Inform. Assoc.* **14** (2007) 29–40. doi:10.1197/jamia.M2170.
- [24] E. Eschmann, P.E. Beeler, G. Zünd, and J. Blaser, Evaluation of alerts for potassium-increasing drug-drug-interactions, *Stud. Health Technol. Inform.* **192** (2013) 1056.
- [25] M.D. Paterno, S.M. Maviglia, P.N. Gorman, D.L. Seger, E. Yoshida, A.C. Seger, D.W. Bates, and T.K. Gandhi, Tiering drug-drug interaction alerts by severity increases compliance rates, *J. Am. Med. Inform. Assoc. JAMIA*. **16** (2009) 40–46. doi:10.1197/jamia.M2808.
- [26] S. Phansalkar, H. van der Sijs, A.D. Tucker, A.A. Desai, D.S. Bell, J.M. Teich, B. Middleton, and D.W. Bates, Drug-drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records, *J. Am. Med. Inform. Assoc. JAMIA*. **20** (2013) 489–493. doi:10.1136/amiainjnl-2012-001089.
- [27] N.R. Shah, A.C. Seger, D.L. Seger, J.M. Fiskio, G.J. Kuperman, B. Blumenfeld, E.G. Recklet, D.W. Bates, and T.K. Gandhi, Improving override rates for computerized prescribing alerts in ambulatory care, *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* (2005) 1110.
- [28] H.R. Strasberg, A. Chan, and S.J. Sklar, Inter-rater agreement among physicians on the clinical significance of drug-drug interactions, *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* **2013** (2013) 1325–1328.
- [29] E.K. Lee, A.F. Mejia, T. Senior, and J. Jose, Improving Patient Safety through Medical Alert Management: An Automated Decision Tool to Reduce Alert Fatigue, *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* **2010** (2010) 417–421.
- [30] E. Chazard, S. BERNONVILLE, G. FICHEUR, and R. BEUSCART, A statistics-based approach of contextualization for adverse drug events detection and prevention, *Stud. Health Technol. Inform.* **180** (2012) 766–770.
- [31] E. Eschmann, P.E. Beeler, V. Kaplan, M. Schneemann, G. Zünd, and J. Blaser, Clinical decision support for monitoring drug-drug-interactions and potassium-increasing drug combinations: need for specific alerts, *Stud. Health Technol. Inform.* **180** (2012) 1200–1202.
- [32] A.M. Miller, M.S. Boro, N.E. Korman, and J.B. Davoren, Provider and pharmacist responses to warfarin drug-drug interaction alerts: a study of healthcare downstream of CPOE alerts, *J. Am. Med. Inform. Assoc. JAMIA*. **18** Suppl 1 (2011) i45–50. doi:10.1136/amiainjnl-2011-000262.
- [33] E. Chazard, G. FICHEUR, S. BERNONVILLE, M. LUYCKX, and R. BEUSCART, Data mining to generate adverse drug events detection rules, *IEEE Trans. Inf. Technol. Biomed. Publ. IEEE Eng. Med. Biol. Soc.* **15** (2011) 823–830. doi:10.1109/TITB.2011.2165727.
- [34] V. Koutkias, V. Kilintzis, G. Stalidis, K. Lazou, C. Collyda, E. Chazard, P. McNair, R. Beuscart, and N. Maglaveras, Constructing Clinical Decision Support Systems for Adverse Drug Event Prevention: A Knowledge-based Approach, *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* **2010** (2010) 402–406.

Tackling Polypharmacy: A Multi-Source Decision Support System

Guilherme REDEKER^{a1} and Juliana BOWLES^a

^a*School of Computer Science, University of St Andrews, St Andrews KY16 9SX, UK*

Abstract. Managing the use of multiple medicines, also known as polypharmacy, is a challenge for physicians, pharmacists and patients alike, and is a particular concern for patients with multiple chronic conditions (aka multimorbidity). Patients with multimorbidity are often required to take a considerable number of medications for their different ongoing conditions, and managing/revising these medications effectively is a challenge. There is a need to periodically rearrange drugs taking into account patient's preferences and avoiding adverse drug reactions. We present an incremental, constraint solver based framework for a clinical decision support system that makes it possible to check drug prescriptions using information from multiple sources, including a constraint database and patient records. We illustrate how it can be used to manage clinical conditions while reducing polypharmacy problems and undesired side effects in a patient-centric approach.

Keywords. Polypharmacy, multimorbidity, constraint solvers, clinical decision support systems

1. Introduction

Polypharmacy can be defined as the concomitant use of multiple medications by a patient [1] [2], and is commonly linked to the treatment of multiple chronic diseases, also known as *multimorbidity*, which occurs with the highest prevalence in the elderly [3]. In addition, the use of many drugs simultaneously to treat different diseases is not always clear and may not match patient expectations and preferences [3] [4]. It is important that the use of various medications is safe and effective [5], and this may become questionable when the risks of medication combinations outweigh the benefits.

Studies have indicated that there are risks associated with the use of multiple medications, which include drug interactions and adverse drug reactions [6] [7]. Drug interactions have a negative outcome when the reaction between drugs has, for instance, a toxic effect or decreases the effect of one drug [6]. An adverse drug reaction is defined as any undesirable medical occurrence caused by a pharmaceutical product but not necessarily related to treatment [8] and may occur for different reasons, such as the wrong dose, drug or route [9]. Studies have shown that over half of the hospital admissions and between 25% to 50% of outpatient care arise from adverse drug reactions which could have been mitigated [10].

The complexity of taking decisions when treating patients can be a significant challenge, especially in the presence of polypharmacy which invariably combine many

¹ Guilherme Redeker, E-mail: gr60@st-andrews.ac.uk.

factors/variables. Additionally, treatments should be personalised and the patient's individual preferences included in the decision process. The position we take is that, specially in the case of complex decision making, automated tool support is essential. Our approach combines medical knowledge with formal methods techniques to search for and compute optimal solutions for managing the treatment of patients with complex needs. This paper is structured as follows: we describe the proposed framework in Section 2 and illustrate how our approach works in Section 3. We conclude the paper with some overall comments and future work in Section 4.

2. Proposed Framework

To support physicians in making appropriate decisions, we propose a clinical decision system framework to check and manage constraints between drugs, exams, and diseases and take into account patient preferences. The solution consists of five parts: the Electronic Medical Record (EMR) interface, the Knowledge interface, the Knowledge database, the Inference engine and the Decision maker

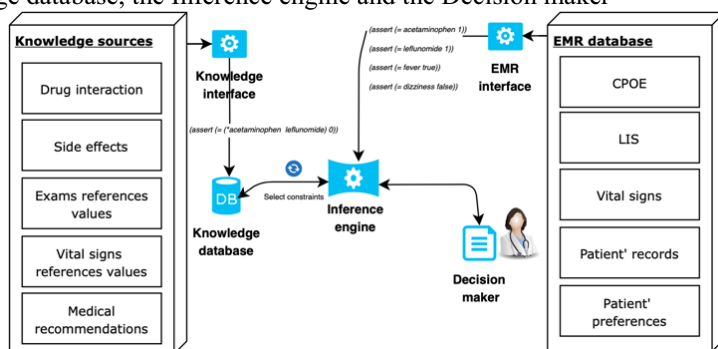


Figure 1. Multi-Source Decision Support System Framework.

The framework shown in Figure 1 takes constraint data from the knowledge sources and patient data from the EMR which are processed independently. The information from both sources is used in the Inference engine, which is responsible for merging both patient and constraint data. The Inference engine checks in the knowledge database if there are constraints that need to be solved according to the patient information then automatically searches for alternative solutions without constraints and displays them for the doctor to decide the appropriate prescription.

2.1. Knowledge interface and database

There are different sources of constraints in knowledge sources which can be used in the Inference engine. To standardize the link between the Inference engine and its sources, we propose a generic knowledge interface. This interface enables the retrieval of data about different constraint types such as drug interactions, side effects, alternative drugs, exam and vital sign reference values. Developing a generic interface will allow users to connect new and different types of constraint sources.

The knowledge interface operates in three steps: receive, process/convert and deliver. The data can be *received* from different channels. Each channel is for a specific type of constraint (e.g., for drug interaction data) and has a custom layout. The data

received by the hub is *converted* into a formal model illustrated in Figure 1 on the *assert* text between the knowledge interface and database, which is compatible with the Z3 [11] solver. *Delivery* is divided into *checking* (searching the knowledge database to see if a specific constraint is already recorded) and *inserting*. If a constraint is already recorded it is discarded, otherwise the system inserts it in the knowledge database to be used by the Inference engine.

2.2. EMR interface

The EMR interface deals with different types of patient data to link the information between the EMR database and Inference engine. Initially, the EMR interface was developed to receive all available information from the patient such as Computerized Physician Order Entry (CPOE) which includes drug prescriptions, laboratory results from the Laboratory Information System (LIS), Vital Signs and Patient Records. The EMR interface also has a channel to receive information on patient preferences. For example, there are medications with side effects such as headache, dizziness or drowsiness, that patients may want to avoid. In these situations, physicians should select drugs which avoid particular side effects.

Similar to the knowledge interface, after receiving information from the EMR, the EMR interface processes the data in three steps: identify, process/convert and deliver. In the first step, the hub *identifies* the type of data (e.g., exam results, drug prescriptions, vital signs or patient records). In the next step, the hub *converts* the data into a formal showed in Figure model to be compatible with the Z3 [11] solver, showed in Figure 1 between the EMR database and interface. Finally, the information is *delivered* to the Inference engine.

2.3. Inference engine and decision maker

The Inference engine was proposed to create links between the knowledge database and patient data, so that it is possible for example to search for drugs that do not have undesired side effects. The compiler receives patients' clinical information and selects the related constraints on the knowledge database to define solutions using the constraint solver Z3 [11].

Firstly, the compiler receives the data from the EMR interface and checks if there are constraints that need to be solved. Secondly, after receiving patient data, a query is executed in the knowledge database based on the received data (e.g., Acetaminophen and Leflunomide drug prescription), to find the linked constraint (in this example, the drug-drug interaction, drug-adverse reaction, drug-disease interaction and drug-food interaction). Finally, if constraints are found for the patient data, the compiler is executed to solve or manage them. As a result, the compiler returns advice to physicians, the decision maker, which takes into account a number of important variables and allows the physician to choose the most appropriate treatment plan to be adopted for a given patient.

3. Practical example

To illustrate how our approach works, we present an extended hypothetical example taken from [5] about a 69 years old man. The patient arrives in hospital after a fainting

and seizures episode and complaining about arm pain and dizziness. The physician checks the patient's previous data on the EMR as shown on the first row of Table 1. Thereafter the physician requests a blood glucose test, verifies the blood pressure and prescribes Dextrose 10% and Electrolyte 500 ml and Codeine phosphate (7,5 mg) + Acetaminophen (300 mg) as shown in the second row of Table 1.

During the consultation, the patient data (previous and new records) is sent from the EMR to the EMR interface and sent to the Inference engine.

Table 1. Patient data

Current medical history	Result exams	Drugs	Patient' preferences
(Previous) Frequent falls Dementia – mixed Alzheimer's disease /alcohol abuse	(Previous) BP 120/74 mmHg Blood glucose test 90 mg/dL	(Continuous use) Trazodone 150 mg Thiamine 50 mg Bendroflumethiazide 2.5 mg	(Previous) Patient reports: feeling tired and short of breath
(New records) Fainting episode, seizures	(New records) BP 132/74 mmHg Blood glucose test 125 mg/dL	(New records) Codeine phosphate (7,5 mg) + Acetaminophen (300 mg) Dextrose 10% and Electrolyte 500 ml	(New records) Patient reports: Feeling arm pain and dizziness

Given the drug interactions and medical recommendations from the knowledge source, the Inference engine checks the constraints from the knowledge database against the patient data and highlights the relevant constraints as in the three tables below.

Table 2. Constraint drug interaction data

Drug	Drug	Interaction	Severity
Trazodone	Codeine phosphate	risk of a rare but serious condition called the serotonin syndrome, which may include symptoms such as confusion, hallucinations, seizures, blurred vision	Moderate
Trazodone	Bendroflumethiazide	lowering blood pressure	Moderate

Table 3. Medical recommendations

Drug	Condition	Recommendation
Dextrose 10%	blood glucose test 125 mg/dL	May cause hyperglycemia

Table 4. Side effects

Drug	Side effects	Used for the following conditions (associated conditions)
Trazodone 10%	dizziness	Alcohol Dependence, Alzheimer's Disease (AD), Dementias
Bendroflumethiazide	dizziness	High Blood Pressure
Codeine	dizziness	Pain, Acute, Severe Pain, Moderate Pain

Thereafter, the physician interactively manages these constraints with the Inference engine. To solve the first interaction in Table 2 between Trazodone and Codeine phosphate, the system shows to the physician the drug interaction and lists alternative drugs to manage the problem. The alternative drugs are selected from the knowledge database according to the associated conditions shown in Table 4. For all listed drugs, the system shows all the possible constraints, for example, interactions with other drugs or side effects that should be avoided according to patient preferences. The alternative drugs are sorted from the best to the more intricate options, that is, from the smallest to the biggest number of related constraints. The physician can also decide

to exclude a drug from the prescription or prescribe a drug that is not listed, for example, to minimise a side effect.

Moreover, if the physician decides to change/include/exclude a drug, the system reloads in real-time the constraints based on this decision. Otherwise, the system keeps the existing constraints and flags the chosen drugs in order to avoid repeated alerts in case of reloading. The same process is repeated to solve the medical recommendations in Table 3 and the side effects in Table 4. The process ends when there are no constraints to be solved, which does not mean that all restrictions have been solved, but that all decisions regarding the constraints have been made.

4. Conclusion

In this paper, we presented the vision of a framework to support physicians to make the appropriate decisions considering different types of drug constraints and interactions, and respecting patient preferences. The framework is enabled to receive multiple data sources combining patient and constraint records. As a clinical decision support system, our proposed automated solution only offers advice and the final decision remains with the physician. We demonstrated how the system can give advice to manage drug conflicts for a realistic but hypothetical example, highlighting the importance of gathering medical evidence and knowledge in decision making. In future work, we intend to fully develop, implement and evaluate the proposed solution.

References

- [1] Cherubini, A., Corsonello, A., & Lattanzio, F. *Polypharmacy in nursing home residents: What is the way forward?* s.l. : Journal of the American Medical Directors Association, 17(1), 4-6., 2016.
- [2] Montero-Odasso, M., Sarquis-Adamson, Y., Song, H. Y., Bray, N. W., Pieruccini-Faria, F., & Speechley, M. *Polypharmacy, Gait Performance, and Falls in Community-Dwelling Older Adults. Results from the Gait and Brain Study.* s.l. : Journal of the American Geriatrics Society, 2019.
- [3] Vasilevskis, E. E., Shah, A. S., Hollingsworth, E. K., Shotwell, M. S., Mixon, A. S., Bell, S. P., ... & Simmons, S. F. *A patient-centered deprescribing intervention for hospitalized older patients with polypharmacy: rationale and design of the Shed-MEDS randomized controlled trial.* s.l. : BMC health services research, 19(1), 165., 2019.
- [4] Curtin, D., Dukelow, T., James, K., O'Donnell, D., O'Mahony, D., & Gallagher, P. *Deprescribing in multi-morbid older people with polypharmacy: agreement between STOPP/Frail explicit criteria and gold standard deprescribing using 100 standardized clinical cases.* s.l. : European journal of clinical pharmacology, 75(3), 427-432., 2019.
- [5] Scottish Government Polypharmacy Model of Care. *Polypharmacy guidance, realistic prescribing.* 2018.
- [6] Brunden, M. N., Vidmar, T. J., & McKean, J. W. *Drug Interaction & Lethality Analysis.* s.l. : CRC Press., 2019.
- [7] Muth, C., Blom, J. W., Smith, S. M., Johnell, K., Gonzalez-Gonzalez, A. I., Nguyen, T. S., ... & Valderas, J. M. *Evidence supporting the best clinical management of patients with multimorbidity and polypharmacy: a systematic guideline review and expert consensus.* s.l. : Journal of Internal Medicine, 285(3), 272-288., 2019.
- [8] Phillips, R., Hazell, L., Sauzet, O., & Cornelius, V. *Analysis and reporting of adverse events in randomised controlled trials: a review.* s.l. : BMJ open, 9(2), e024537., 2019.
- [9] Cornell, E., Kwa, M., Paller, A. S., & Xu, S. *Adverse events reported to the Food and Drug Administration from 2004 to 2016 for cosmetics and personal care products marketed to newborns and infants.* s.l. : Pediatric dermatology, 35(2), 225-229., 2018.
- [10] Cahir, C., Wallace, E., Cummins, A., Teljeur, C., Byrne, C., Bennett, K., & Fahey, T. *Identifying Adverse Drug Events in Older Community-Dwelling Patients.* s.l. : The Annals of Family Medicine, 17(2), 133-140., 2019.
- [11] Moura LD, Bjørner N. Z3: an efficient SMT solver. In: TACAS 2008. Lecture Notes in Computer Science, vol 4963. Springer, Berlin, pp 337-340, 2008

The House of Carbs: Personalized Carbohydrate Dispenser for People with Diabetes

Pietro RANDINE^{a,b,1}, Daniela MICUCCI^c, Gunnar HARTVIGSEN^a, Eirik ÅRSAND^{a,b}

^a*Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway*

^b*Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway*

^c*Department of Informatics, Systems and Communication (DISCo), University of Milano Bicocca, Milan, Italy*

Abstract. Patients with diabetes are often worried about having low blood glucose because of the unpleasant feeling and possible dangerous situations this can lead to. This can make patients consume more carbohydrates than necessary. Ad-hoc carbohydrate estimation and dosing by the patients can be unreliable and may produce unwanted periods of high blood glucose. In this paper we present a system that automatically estimates and dispenses the amount of juice (or similar) according to the current patients' blood glucose values. The system is remotely accessible and customizable from a chatbot, exploits sensors and actuators to dispense the necessary amount of liquid carbohydrates. It relies on a cloud solution (Nightscout) to acquire the patient's blood glucose values, which are constantly updated thanks to a commercial wearable continuous glucose monitor (CGM).

Keywords. Diabetes, Hypoglycemia, Carbohydrates, Cyber-Physical Systems, Internet of Things

1. Introduction

Measuring and managing blood glucose concentration is a key element in diabetes care, especially in type 1 diabetes. Since the first real time continuous glucose monitor (CGM) was provided for patients in 2001 [1], its relevance in both ambulatory diabetes care and clinical research has increased over the years [2]. Decreasing size, weight, complexity and cost of CGM sensors/devices have increased usage and dissemination [3]. CGMs have helped patients to improve their quality of life, by increasing their confidence in self-management of their disease.

In diseases such as diabetes, due to the availability of low cost technologies, well-educated patients, or engaged relatives, it has become possible to formulate, develop and distribute solutions that aim to answer specific needs in managing the disease, based on the patients individual situation [4].

¹ Corresponding author: email: pietro.randine@uit.no

In this study we show how it is possible to integrate patient-initiated solutions such as Nightscout [5], different sensors, mobile technologies and actuators, as concrete tools in patients’ daily life. The system referred to as “The House of Carbs”, was developed in collaboration with patients, to help them in one of the daily challenges in insulin-treated diabetes, namely low blood glucose episodes (hypoglycemia). The goal was to design a solution that can automatically dose an estimated amount of juice (or similar) from a reservoir to a glass, when required. The motivation is to avoid serious hypoglycemia situations and in preventing subsequent high levels of blood glucose (hyperglycemia) due to intake of more carbohydrates than necessary. We aim in stabilizing the blood glucose and avoid the “Roller coaster effect”, when the blood glucose is fluctuating between levels outside the normal range.

This paper is organized as follows: first we describe the methods used, and then we present the designed prototype and data from 2 months of use by two patients. The paper concludes with limitations and challenges for integrating sensors and actuators in a cyber-physical system for people with type 1 diabetes.

2. Methods

A literature review was conducted for mapping the available systems for managing hypoglycemia. The outcomes of this review and a patient-involvement approach were then used to define the system functionalities and requirements. Patients’ needs and expectations were also used to improve the software during the prototype testing phase.

Figure 1 shows the conceptual overview of the system, and how it interacts with external hardware and software components.

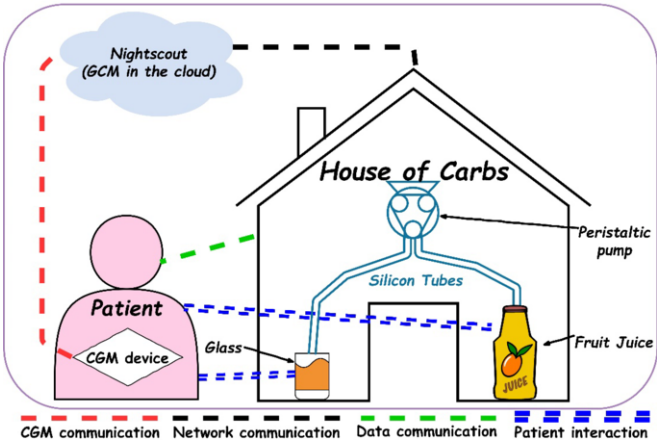


Figure 1: House of Carbs – a conceptual overview

Nightscout provides the real time data necessary to detect the hypoglycemic episodes and to evaluate the result achieved by the liquid carbohydrates delivery.

The patient must be an active part of the system, (s)he will have to insert the silicon tube inside the juice container and glass. Furthermore, the patient needs to provide the information about the juice’s content of carbohydrates and the web address of the Nightscout platform. In addition, (s)he is responsible for proper maintenance and use of the system, e.g. if the glass is removed from the system, the consequent assumption is

that the patient has drunk its content, and it needs to be replaced with a new glass before the next delivery.

3. Results

3.1. The prototype “House of carbs”

We built a fully functioning prototype using a Raspberry Pi 3 and an add-on board to control two high-power DC motors. The “House of carbs” includes an additional chatbot application to remotely control the hardware (e.g., clear the pump, remotely turn off the machine), define system parameters (e.g., carbohydrate content and amount of juice present, Nightscout address, units), and send notifications (e.g., “dose ready to be picked up”, “glass removed”), see Figure 2.

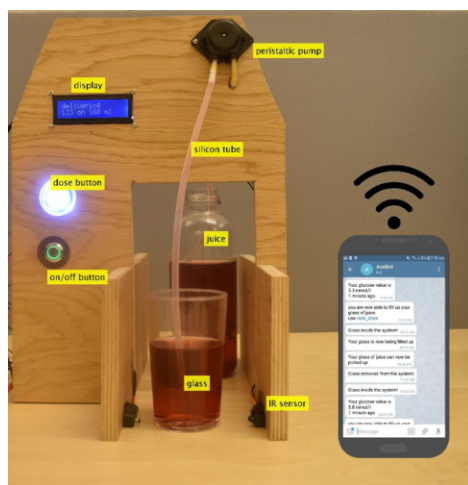


Figure 2: The prototype “House of Carbs”, based on communication with Dexcom G4 CGM and the Nightscout platform.

The system reports its status and the blood glucose values through a LCD display. The detection of the glass represents a critical issue in this system, and this is done by IR sensor. In case the glass is not detected, the system will block any juice distribution and users will receive a notification about the missing glass via the chatbot.

3.2. Treatment of hypoglycemia and prevention

A common approach to treat hypoglycemia is the 15/15 rule [6] (eat 15 grams of carbohydrate and wait 15 minutes), especially effective when using liquid carbohydrate sources [7]. This rule suggests checking the blood glucose value after 15 minutes and repeating the procedure if the glucose value is still under 3.9 mmol/L (70 mg/dL). The proposed system was designed together with real users, revealing the importance of automatically detect hypoglycemic episodes and deliver carbohydrates. The amount of carbohydrates was made personalizable, and the users can choose the amount of delivered carbohydrates (e.g. 20g instead of 15g). According to their suggestion, the system allows patient to request carbohydrates when expecting hypoglycemia, especially

relevant before an intense activity (e.g., physical activity) [8] or for safety reason before activities that may be risky (e.g. driving a car) [1].

3.3. Testing

The trial ran for two months involving two persons with type 1 diabetes. During this period the system logged 3116 different CGM readings and distributed 52 doses of carbohydrates. The system delivered either 15 or 20 grams of carbohydrates each time, based on the patient’s preferences. Table 1 presents the summary of these distributions.

Table 1: Successful juice distribution divided by patient and dose type (hypoglycemic or request)

Patient no	Carbs (g)	Total doses	Description	Value
1	15	9	Average hypoglycemic glucose value detected before the dose (mean ± SD)	3.48 ± 0.51 mmol/L
			Average glucose value after approx. 15 minutes (mean ± SD)	4.17 ± 0.79 mmol/L
	20	20	Average hypoglycemic glucose value detected before the dose (mean ± SD)	3.56 ± 0.43 mmol/L
			Average glucose value after approx. 15 minutes (mean ± SD)	4.13 ± 1.08 mmol/L
2	20	2	Average hypoglycemic glucose value detected before the dose (mean ± SD)	3.9 ± 0 mmol/L
			Average glucose value detected after the dose (mean ± SD)	4.7 ± 0 mmol/L
1	15	6	Average glucose value detected before dose request (mean ± SD)	4.75± 0.59 mmol/L
			Average glucose value after approx. 20 minutes (mean ± SD)	4.17 ± 1.38 mmol/L
	20	15	Average glucose value detected before dose request (mean ± SD)	4.85 ± 0.62 mmol/L
			Average glucose value after approx. 19 minutes (mean ± SD)	4.77 ± 0.83 mmol/L

4. Discussion

Health systems involving sensors and actuators need to be robust and adaptable to unexpected conditions and subsystem failures [9]. However, no component is perfectly reliable, and the physical environment will produce unexpected conditions.

The presented system integrates the open source diabetes-monitoring tool Nightscout, that has an informational and educational purposes only. Thus, the CGM data from Nightscout were the unique source of information in this system – which only was made to present future possibilities. All the designed functionalities are dependent

of this data, and when no data are received from Nightscout, the system is not able to provide these functionalities. Additionally, to maximize the accuracy of the system, the sensors connected to Nightscout must be calibrated according to the device producer's instructions. It is important to make it clear that the test period was short, the number of subjects were limited, and the patient participants have more than 20 years of diabetes experience and involved in different diabetes research projects.

4.1. Practical limitations

Designing such a system requires several assumptions. For example, that the source of carbohydrates is a safe and efficient liquid, the target glass has a capacity that can hold the maximum delivered dose, and both the source and target containers need to be clean due to hygienic reasons. Also, the entered settings need to be accurate for the system to deliver the right doses. If the user enters incorrect information (e.g., the carbohydrates content of the juice), the system may not help the user to achieve a healthy blood glucose level. Furthermore, the chatbot component of the system can potentially be accessed by anyone and therefore additional security solutions is needed to avoid abuse or misuse.

5. Conclusion

The main contribution of this research is demonstrating the design of a practical system able to monitor and prevent hypo- and hyperglycemia. The system has been designed using a patient driven approach, and integrates some already existing components.

The absorption of glucose into the bloodstream was shown to be effective for the two test-persons, after one or more delivered doses, in average after 15-20 minutes. They were satisfied with the use of the system and even surprised how it was able to stabilize the blood glucose. No hyperglycemia episodes or "Roller coaster effect" were caused by the system.

References

- [1] Bailey, T.S., et al., American Association of Clinical Endocrinologists and American College of Endocrinology 2016 outpatient glucose monitoring consensus statement. *Endocrine Practice*, 2016. 22(2): p. 231-261.
- [2] Schnell, O., et al., Role of continuous glucose monitoring in clinical trials: recommendations on reporting. *Diabetes technology & therapeutics*, 2017. 19(7): p. 391-399.
- [3] Rodbard, D., Continuous glucose monitoring: a review of successes, challenges, and opportunities. *Diabetes technology & therapeutics*, 2016. 18(S2): p. S2-3-S2-13.
- [4] Omer, T., Empowered citizen 'health hackers' who are not waiting. *BMC Med*, 2016. 14(1): p. 118.
- [5] Project, N., NIGHTSCOUT# WeAreNotWaiting. 2017, Nightscout Project McKinney, TX.
- [6] Tomky, D., Detection, prevention, and treatment of hypoglycemia in the hospital. *Diabetes Spectrum*, 2005. 18(1): p. 39-44.
- [7] Kreider, K.E., B.I. Padilla, and K. Pereira, Hypoglycemia in diabetes: challenges and opportunities in care. *The Journal for Nurse Practitioners*, 2017. 13(3): p. 228-234.
- [8] Seaquist, E.R., et al., Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *The Journal of Clinical Endocrinology & Metabolism*, 2013. 98(5): p. 1845-1859.
- [9] Lee, E.A. Cyber physical systems: Design challenges. in 2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC). 2008. IEEE. Fossum M, Ehnfors M, Svensson E, Hansen LM, Ehrenberg A. Effects of a computerized decision support system on care planning for pressure ulcers and malnutrition in nursing homes: An intervention study. *Int J Med Inform* 82/10 (2013), 911-921.

Update on the DMP, the French Nationally Shared Medical Record: Did We Make It?

Brigitte SÉROUSSI^{a,b,1} and Jacques BOUAUD^{c,a}

^a*Sorbonne Université, INSERM, Université Paris 13, Sorbonne Paris Cité, UMR_S
1142, LIMICS, Paris, France*

^b*AP-HP, Tenon Hospital, Paris, France*

^c*AP-HP, DRCI, Paris, France*

Abstract. Different countries have solved the problem of care coordination by implementing a nationwide centralized framework of clinical information sharing with “new” secure online care records stored in specifically created platforms. The French DMP initially launched in 2006, relaunched in 2010, and re-relaunched in 2016 follows this model. After a difficult start and some governmental actions to promote its adoption, the DMP has been nationally deployed in November 2018. New features have been developed (creation of DMPs for children, management of minors’ rights, addition of living wills, notification of the first access of any healthcare professional), and financial incentive measures have been implemented. In September 2019, over seven million people had a DMP in France. In 18 months, we observed a significant increase of GPs feeding (from 0.7% to 20.9%) and consulting (from 1.6% to 44.8%) DMPs.

Keywords. Personal Health Records, Electronic Health Records/utilization, Patient Access to Records, Patient Participation, Mobile application

1. Introduction

Patients suffering of chronic diseases and related morbidities are currently managed in multiple settings by numerous healthcare professionals. Care is thus currently strongly fragmented and care coordination among the different healthcare professionals is complex. As a result, care management may be not effective and sometimes harmful to the patient [1]. The potential of electronic health records (EHRs) and other health information technology (HIT) tools to facilitate information flow at care transitions has been described [2]. In particular, these tools allow the different care providers to share the same information about a patient’s condition which allows to access crucial information in emergency room and optimizes the quality of care, reduce adverse drug reactions considered as the leading causes of morbidity and mortality in health care, and decreases health care costs by improving care relevance and not duplicating diagnostic tests. However, while EHRs are necessary, they are not sufficient to guarantee care coordination because non-interoperable EHRs and vendors’ expensive custom interfaces are hindering health information exchanges [3].

¹ Corresponding Author, brigitte.seroussi@aphp.fr.

Different countries have implemented a nationwide centralized framework of clinical information sharing to solve the problem of care coordination relying on the development of “new” secure online care records stored in specifically created platforms [4]. Adoption of these medical records has historically been poor and governmental actions were taken to promote their uptake. For instance, the DMP or “Dossier Médical Personnel”, currently “Dossier Médical Partagé”, has been relaunched in France in 2010 after a first start in 2006 and a first stop six months later due to security concerns [5]. It has also been re-launched in 2016 because of the poor uptake after the first relaunching (only 520,000 DMPs and about a coverage of 1.5% of the target population). At this time, the DMP program governance was transferred to the French national health insurance or CNAM (instead of the national technical agency in charge of HIT systems interoperability or ASIP Santé) and acted by the 2016 French Health Act.

Since 2016, new functionalities have been developed to improve the uptake of the “new” DMP, including the creation of DMPs by patients themselves, the automatic post of health data derived from insurance claims at the creation of DMPs (giving some medical contents to DMPs), the publication of a mobile app, and the organization of large public communication campaigns to inform patients and more widely involve citizens. These new features have been tested between December 2016 and August 2018 in nine pilot French departments with a good success [6]. This paper describes the evolution of the DMP program since August 2018, and specifically after the DMP official national deployment in November 2018 by both the French Ministry of Health and the French national health insurance’s director.

2. Methods

Originally, the DMP project was designed to support care coordination. All care providers may populate their patients’ DMP (discharge summaries, lab results, radiology reports, etc.). However, only care providers authorized by a DMP’ owner may access and read his/her DMP contents. To promote the national deployment of the DMP program, new functionalities have been developed to improve the quality of services provided by DMPs, along with the implementation of financial incentive measures varying with the different types of healthcare providers.

First, a huge work has been done by the French national health insurance with EHR vendors and care professionals to resolve remaining technical issues and promote the smooth interoperability of the DMP and EHR systems (almost all private physicians and hospitals have been accompanied by French national health insurance technical agents to help them upgrading their EHR systems). In addition, new features have been implemented such as the creation of DMPs for children, the management of minors’ rights (although under the responsibility of their legal representatives who are in charge of managing their DMP, minors may hide information to their parents with the support of an adult and in relationship with their general practitioner, GP), and the creation of a DMP chapter to collect advance directives (living wills). Besides, since DMPs are technically open to all healthcare professionals (except to those that have been explicitly excluded by the patient) to allow emergency accesses by any healthcare professional if necessary, the patient owner of a DMP is now notified of the first access of any healthcare professional to his/her DMP, in addition of this access being tracked in the DMP journal of accesses (only accessible to the patient and his/her GP). Finally,

a second version of the DMP app has been developed to improve interface ergonomics and usability.

Financial incentives have been implemented to accompany the deployment of the DMP program. Because pharmacists were very good to create the patient pharmacological record (DP), the French national health insurance offered them one euro per DMP created (nurses are the next on the list to get this incentive). DMP compatibility with EHRs used by private practitioners is now a prerequisite to get the ROSP (“Rémunération sur Objectifs de Santé Publique) that accounts to a maximum of 3500 euros per year. Biologists are also encouraged to use DMP-compatible HIT tools and they may receive from 5,500 to 10,000 euros per year (according to their levels of activity) when pushing biological reports / lab results in DMPs. Hospitals and private clinics may consider entering the nationwide HOP’EN program (HOP’EN builds on healthcare institutions digital transformation and modernization, the aim being that by 2022, whatever their status, size, and activity, they could bring their HIS to the maturity level necessary to address the new challenges of the health system defragmentation) and get revenues when they post discharge summaries in their patients’ DMP. Finally, nursing homes are also supported to use a DMP-compatible information system and to push post emergency liaison records in DMPs (they can receive up to 2,500 euros).

3. Results

Whereas there were 580,000 DMPs in June 2016, and 930,000 DMPs by the end of October 2017, there were seven million DMPs on September 15th, 2019. Currently, 10,000 DMPs are created every week. Figure 1 displays the evolution of DMP creation over the time. The acceleration of DMP creation after the national deployment of the program is visible. DMPs are mostly created in French national health insurance offices (45%), pharmacies (35%), and by patients using the DMP website on Internet.

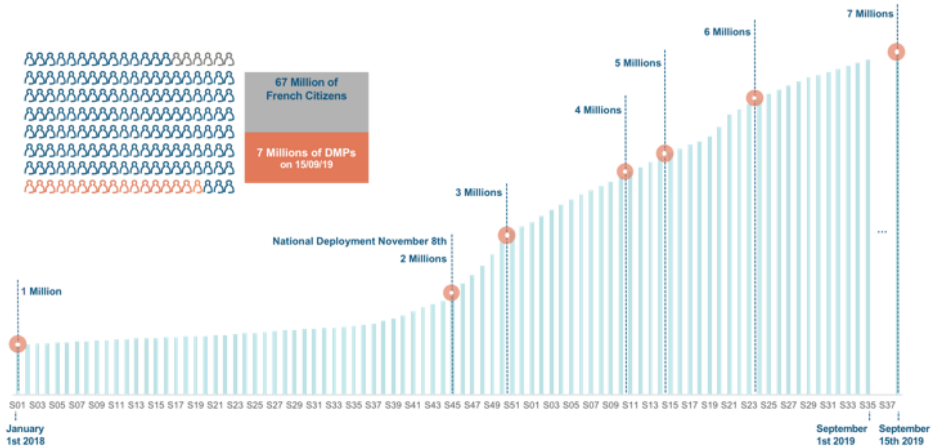


Figure 1. Evolution of DMP creation visualizing the acceleration of the process after the national deployment of the DMP program.

DMP contents essentially come from private clinics and hospitals inputs. Currently, 400 out of the 3000 French care structures are pushing discharge summaries in DMPs (24 university hospitals out of a total of 32 nationwide). Besides, there is a significant increase of DMP feeding by GPs (from 0.7% to 20.9% in 18 months), and a significant increase of DMP consultation by doctors (from 1.6% to 44.8% for GPs, and from 0.1% to 7.3% for other medical specialists). The case of pharmacists is also noticeable with an increase of DMP consultation from 0.0% to 11.3%. Currently, and for political reasons (there is still no consensus between the French national health insurance and biologist unions on how, when, and which levels of incentives should be provided), biologists remain less involved in DMP creation, population, and consultation. Table 1 displays the evolution of DMP population and consultation according to the type of care practitioners.

Table 1. Evolution of DMP population and consultation according to practitioners types in 18 months.

Type of care practitioners	Number of care practitioners nationwide	January 2016 - December 2016				September 2018 - August 2019			
		Populating DMPs	%	Consulting DMPs	%	Populating DMPs	%	Consulting DMPs	%
General Practitioner	52 641	383	0,7%	840	1,6%	10 996	20,9%	23 567	44,8%
Medical Specialist	54 555	35	0,1%	56	0,1%	1 629	3,0%	3 993	7,3%
Masseur-Physiotherapist	67 860	19	0,0%	39	0,1%	98	0,1%	1 126	1,7%
Nurses	93 206	33	0,0%	126	0,1%	132	0,1%	815	0,9%
Other professions	77 562	3	0,0%	11	0,0%	144	0,2%	1 130	1,5%
Medical Biology Laboratories	3 864	6	0,2%	5	0,1%	12	0,3%	10	0,3%
Pharmacies	21 510	1	0,0%	3	0,0%	542	2,5%	2 437	11,3%

A new DMP mobile app has been developed to improve the display of all the documents contained in the DMP. A totally updated interface has been implemented, starting with a home page displaying the last documents added as well as the identity of the last care provider that accessed to the DMP. New features have been developed such as the access to images, and a button to help the patient hide some information of his/her DMP. Some screen shots of the new DMP mobile app are given in Figure 2.

4. Discussion and conclusion

After a difficult starting period, the French DMP looks to be successfully adopted by both patients and care providers. However, although the pace of DMP creation is impressive, DMP feeding is substantially increasing, and DMP consultation is beginning to happen, the DMP remains unnatural for many care providers that consider it as an extra tool on top of their usual EHR. Thus, new features are currently under development to make DMPs attractive, such as a chapter to represent the vaccination record, and the connection with mobile devices to collect relevant medical data, but it is not clear whether this will be enough and patients who are active supporters of the DMP program have to ask their care providers to populate and consult their DMP.

The next step will be the switch from the opt-in to the opt-out model where DMPs will be automatically created unless the patient opposed to it. This has been written in the last French Health Law (2019) and should be performed by the beginning of 2022.

This is similar to what happened in Australia with “My Health Record” and it is when the program started to encounter serious difficulties [7] because of public distrust over data privacy. We hope the transition to be smoother in France and expect about 60 million DMPs by 2022 (to account for the usual 10% of opt out rate). At this time, we would know if we really made it.

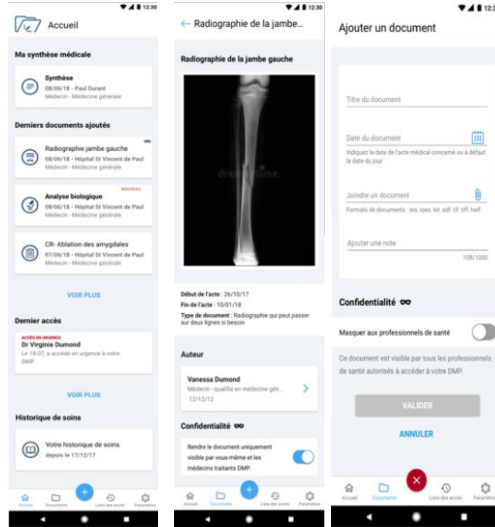


Figure 2. Screen shots from the new DMP mobile app.

Acknowledgements

Authors thank Yvon Merlière, project leader of the DMP program at the CNAM, for the data and figures on the DMP creation, and for the screenshots from the mobile app.

References

- [1] T. Bodenheimer. Coordinating care: a perilous journey through the health care system. *N Engl J Med* **358**/10 (2008), 1064–1071.
- [2] I. Graetz, M. Reed, T. Rundall, J. Bellows, R. Brand, J. Hsu. Care Coordination and Electronic Health Records: Connecting Clinicians. *AMIA Annu Symp Proc* (2009), 208–212.
- [3] R. Parks, RT. Wigand, MB. Othmani, Z. Serhier, O. Bouhaddou. Electronic health records implementation in Morocco: Challenges of silo efforts and recommendations for improvements. *Int J Med Inform* **129** (2019), 430–437.
- [4] S. de Lusignan S and B. Séroussi. A comparison of English and French approaches to providing patients access to Summary Care Records: scope, consent, cost. *Stud Health Technol Inform* **186** (2013), 61–65.
- [5] B. Séroussi and J. Bouaud. Adoption of a Nationwide Shared Medical Record in France: Lessons Learnt after 5 Years of Deployment. *AMIA Annu Symp Proc* (2016), 1100–1109.
- [6] B. Séroussi, J. Bouaud. The (Re)-Relaunching of the DMP, the French Shared Medical Record: New Features to Improve Uptake and Use. *Stud Health Technol Inform* **247** (2018), 256–260.
- [7] D. Lupton. 'I'd like to think you could trust the government, but I don't really think we can': Australian women's attitudes to and experiences of My Health Record. *Digit Health* **29**/5 (2019), 2055207619847017.

User Requirements Meet Large-Scale EHR Suites: Norwegian Preparations for Epic

Gunnar ELLINGSEN^{a,1} and Morten HERTZUM^b

^a *UIT – The Arctic University of Norway, Tromsø, Norway*

^b *University of Copenhagen, Copenhagen, Denmark*

Abstract. Electronic health records (EHRs) are becoming the norm in healthcare. Typically, these EHRs are large-scale suite systems. The up-front presence of ready-for-use EHR suites changes the role of user requirements and the conditions for deciding which requirements to include in the final contract. In this paper, we investigate how user requirements are negotiated in the ongoing preparations for the implementation of Epic throughout the region of Central Norway. User requirements shape vendor selection but they are also shaped by the vendors' existing EHRs and by the requirements of the selected vendor's other customers.

Keywords. Electronic health record, Epic, implementation, user requirements

1. Introduction

Electronic health records (EHRs) are becoming the norm in modern healthcare. Typically, these EHRs are large-scale suite systems with many modules and extensive configuration facilities [1]. Examples include Cerner, Epic, and InterSystems. In the initial procurement phase healthcare organizations choose among vendors that already have EHR suites ready for deployment. Subsequently, healthcare organizations negotiate configurations, extensions, and revisions with selected vendors. The initial presence of ready-for-use EHR suites changes the role of user requirements and the conditions for deciding which requirements to include in the final contract [2]. We investigate how user requirements are negotiated in the ongoing preparations for the implementation of Epic throughout the region of Central Norway. The new EHR in Central Norway is not restricted to its hospitals but will also cover its municipal healthcare services, that is, general practitioners (GPs), home care, and nursing homes. Due to the heterogeneity of these healthcare providers, the user requirements are many and diverse. It is essential to convey these requirements to the vendors. After a single vendor has been selected it is especially important that this vendor comes to understand the requirements. At the same time the requirements – and the implementation preparations in general – will be strongly shaped by the EHR suite already available from the vendor. The outcome of this process is still open [3].

We ask: *How are user requirements negotiated and shaped in the meeting with large-scale EHR suites?* Empirically, we ask this research question in relation to the Epic implementation in Norway, but the process is not specific to the Norwegian case.

¹ Corresponding Author: Gunnar Ellingsen, E-mail: gunnar.ellingsen@uit.no

2. Method

The study takes an interpretive research approach [4,5]. The data consist of interviews and publicly available documents. We conducted six interviews with top management in the Health Platform program, which is responsible for the Epic implementation in Central Norway. The director of the program was interviewed twice. The interviews were audio-recorded and transcribed for analysis. Media coverage and informal discussions with project members provided supplementary input. The data were collected in the autumn of 2018, about half a year before the contract with Epic was signed.

3. Case

The Health Platform is a regional program owned by the Central Norway Regional Health Authority and Trondheim municipality. It aims to acquire and implement an EHR for the whole region, including all the hospitals, GPs, nursing homes, and home-care services. By including the municipalities, the health authorities want to make the Health Platform a pilot for the long-term national goal of establishing a nationwide EHR functionality, "one citizen - one record" [6].

The Central Norway region includes 44,000 healthcare professionals and an approximate population of 720,000 citizens. There are 3 hospitals whereof the largest is the university hospital, St Olav's Hospital, located in Trondheim. St Olav's Hospital is already at a high level of digitization. In contrast, the level of digitization varies substantially across the GPs, nursing homes, and home-care services, where particularly the GPs already have well-working systems at their disposal. The Health Platform has identified 80 current information systems that will be replaced by the new EHR. In addition, 160 integrations between the EHR and other systems must be developed. Trondheim municipality will be in the first wave of implementation, while the rest of the 84 municipalities in Central Norway have the option to opt in after that.

3.1 *The bid for tender*

The invitation to tender was announced in August 2016 with an upper limit for the contract of NOK 2.7 billion (EUR 270 million), including licenses, the vendor's services, and 10 years of maintenance [7]. The time schedule for the main activities was as follows: vendor prequalification in 2016, dialog phase with selected vendors in 2017/2018, contract with a specific vendor signed in 2019, and implementation and go-live in 2021.

From early on, extensive user participation was identified as crucial to the success of the program. When the acquisition process started, 400 clinicians from municipalities and hospitals across the Central Norway region became involved. These clinicians participated in 101 workshops in which they described what they valued about their current information systems, what challenges they currently experienced, and what was missing. The outcome of these workshops became the basis for the requirements specification, which amounted to 4000 user requirements.

A total of 11 EHR vendors applied for prequalification in 2016. In February 2017, the Health Platform sent out the final information for the procurement of the future HER

to the five prequalified EHR vendors: Cerner, CSC Scandihealth (later renamed to DXC Technology), Epic, InterSystems, and Tieto. The tender was based on the user requirements and a dialogue with the vendors about how they planned to meet these requirements. The bids submitted by the vendors were evaluated by project members and by people from the line organization. In May 2017 four EHR vendors submitted revised bids, and in January 2018 the number of vendors was reduced to two, namely Cerner and Epic. However, Cerner decided to withdraw, because they could not see a business case for making a bid that met all requirements in the functionality for the municipalities. The contract between the Health Platform and Epic was signed in March 2019.

3.2 Negotiating the development of functionality

In spite of the substantial functionality and configurability of Epic, some new software development was needed, particularly to extend Epic with additional functionality for the municipalities. The Health Platform requested that Epic developed a pregnancy record, a medication module for home care, and support for patient-related administrative procedures such as which benefits the patient shall receive. Although Epic initially responded positively to these requests, things became more complicated closer to the deadline for signing the contract. The Health Platform had to work hard to convey Norwegian routines to Epic, particularly in relation to the administrative procedures.

A lot of functionality is dependent on integrations between Epic and other systems. In the municipalities, the systems that need to be integrated with Epic include Transmed at the security centers, employee administrative systems, economy systems, and the roster module in the administrative resource management system. The need for integration invoked questions about the location of certain data and functions. For example, the GPs' current EHR stores their economy data. As a result, all claims against, for example, the Norwegian Labor and Welfare Administration are handled in their EHR. In addition, the GPs have billing systems where patients pay for their GP visit. The vendors of the billing systems handle debt collection from patients who have not paid what they owe according to the data in the GPs' EHR. Epic was not familiar with such a setup, but with one in which the billing system handled all economy data. It requires considerable work to develop the integration between Epic and the billing systems.

To ensure the continued evolution of the EHR after its implementation, the Health Platform management envisions to increase its collaboration with other European Epic customers. This is in line with the Epic strategy for development. As a means to establish common requirements, Epic encourages its customers to collaborate on requirements for new functionality. For example, there is currently a collaboration between Epic customers in Switzerland and Denmark about the development of functionality for medication management. These collaborations mean that new Norwegian user requirements will have to be brought into alignment with those of other customers as well as to be negotiated with Epic. A new version of the entire Epic system is rolled out to all customers four times a year; the philosophy is that all new developments are available to all customers.

3.3 Configuring the system

Overall, the Health Platform management was satisfied with the configuration facilities in the system. The Health Platform was, for example, pleasantly surprised when they

learned that Partogram (graphical maternal and fetal data) functionality could be configured and, thus, did not have to be developed from scratch by Epic. However, Epic promoted configuration proactively, and sometimes at the expense of development. An illustrative example is the Health Platform's request for offline EHR functionality in the ambulances. When ambulance personnel lose online contact with the EHR, they should still be able to read and update the EHR in an offline mode, and these updates should then be synchronized with the online version later. Epic argued that the development of offline functionality would delay the project. As an alternative, they suggested that they could configure the existing user interface of the ambulance module in a way that would partially cover the requested functionality. After conferring with the relevant clinical disciplines, the Health Platform decided to let go of offline functionality and settle for the solution that Epic offered.

Configuring the Epic system in the start-up phase involves lots of decisions about how the system should function. While Epic can be configured differently for different settings (e.g., GPs, home-care services, hospitals, and nursing homes), a great deal of standardization of work processes is necessary to achieve the goal of an all-compassing EHR that can support inter-organizational information exchange and workflows. Standardization can also be a means of getting rid of unwanted variation in treatment and care. For a given patient condition and a given step in, say, an outpatient work process, the system could for example enforce the prescription of certain drugs or tests, and in a specific order. The Health Platform has established an extensive formal decision structure to decide on such standardizations, but some decisions will also require top-level approval because the standardization has far-reaching organizational consequences. For example, the Health Platform is considering to enforce a fully electronic medication management process (ordering, dispatching, distribution, dispensing, and administration), which currently is only supported by paper-based routines.

The formal decision structure will be pertinent to a successful implementation. During the preparations for going live in 2021, it is estimated that Epic will raise something like 8000-12000 questions about the details of how the system should be configured. The customer must respond within a 10-day deadline, otherwise Epic will choose the default configuration for the area in question.

After Epic has been implemented, key users (and not the vendor) will be responsible for the continued configuration of the system. In Epic terminology, these clinicians are called "physician builders". They will become responsible for configuring the layout of the system, its information flows, and its support of workflows. During the configuration of the initial Epic setup the physician builders will work together with Epic personnel. In a subsequent regional organization, the physician builders will continue to work on optimizing and streamlining work processes. To prepare for this task, the physician builders will attend a training course of up to six weeks and will subsequently work full-time as physician builders in Central Norway. This policy is intended to ensure flexibility for the users and to reduce dependence on the vendor.

4. Concluding discussion

In the initial procurement phase of EHR suites, the customer has the option of choosing between different vendors based on a detailed requirement specification. The vendors present what functionality they offer, and the development of new functionality is

negotiated. In this phase, the vendors must be able to respond satisfactorily to the customer's demands. However, after the vendor selection the preparations for implementing the system are strongly shaped by the product already available from the vendor, including its configuration possibilities. During these preparations the customer's (and future users') involvement consists to a large extent of responding to opportunities provided by the vendor, such as to queries about configuration choices. In this process the requirements specification may face considerable competition and initial requirements may be revised on the basis of descriptions and demonstrations of the vendor's suite system [8]. We are not claiming that requirements specification is no longer important but that its main role is in vendor selection.

After go-live the vendor will most likely be less available, both for system development and for configuration. New functionality that requires development must be aligned with other Epic customers along the lines of the current collaboration between Switzerland and Denmark. By encouraging such collaborations, the vendor shifts it onto its customers to reach agreement about how the system should evolve [9]. This way alliances and shared – or conflicting – interests among the customers will shape the user requirements that the Health Platform can present to Epic. Requirements specific to single customers will likely receive low priority from Epic or be considered topics for configuration (by the individual customer) rather than development (by Epic).

Regarding future configuration, it is expected that the physician builders do this. However, given that many decisions about configuration will have to be made before go-live, it is an open question to what extent it will be practically possible for the physician builders to make larger configurations later. After all, some of these decisions must be made in a complex decision structure, sometimes requiring top-level organizational approval. For larger configurations to happen, a continued organizational commitment to the same extent as in the implementation phase will probably be necessary. Without such a continued commitment, the extensive configuration possibilities will remain a dull instrument for meeting emergent user requirements.

References

- [1] A.J. Holmgren, J. Adler-Milstein, J. McCullough, Are all certified EHRs created equal? Assessing the relationship between EHR vendor and hospital meaningful use performance, *Journal of the American Medical Informatics Association* **25** (2018), 654-660.
- [2] G.S. Fletcher, T.H. Payne, Selection and implementation of an electronic health record, *PM&R* **9** (2017), S4-S12.
- [3] M. Hertzum, G. Ellingsen, The implementation of an electronic health record: Comparing preparations for Epic in Norway with experiences from the UK and Denmark, *International Journal of Medical Informatics* **129** (2019), 312-317.
- [4] H. Klein, M. Myers, A set of principles for conducting and evaluating interpretive field studies in information systems, *MIS Quarterly* **23** (1999), 67-94.
- [5] G. Walsham, Interpretive case studies in IS research. Nature and method, *European Journal of Information Systems* **4** (1995), 74-81.
- [6] Direktoratet for e-helse, Utviklingstrekk 2018 - Beskrivelser av Drivere og Trender Relevant for e-helse. Versjon 1.0, Oslo, Norway (2018).
- [7] Helse Midt-Norge-1, Vedtok tildeling av kontrakt om Helseplattformen, <https://helse-midt.no/nyheter/2019/vedtok-tildeling-av-kontrakt-om-helseplattformen> 2019 (accessed Aug. 8, 2019).
- [8] A. Finkelstein, G. Spanoudakis, M. Ryan, Software package requirements and procurement. In *Proceedings of the 8th International Workshop on Software Specification and Design*, Schloss Velen, Germany, London: IEEE Computer Society Press (1996), 1-7.
- [9] N. Pollock, R. Williams, L. D'Adderio, Global software and its provenance: Generification work in the production of organizational software packages, *Social Studies of Science* **37** (2007), 254-280.

Value Mechanisms in the Implementation of Intelligent Patient Flow Management System – A Multiple Case Study

Tuomo TANILA^a, Henni TENHUNEN^{a,1} and Petteri HIRVONEN^b

^a*Institute of Healthcare Engineering, Management and Architecture (HEMA Institute),
Department of Industrial Engineering and Management, Aalto University, Espoo,
Finland*

^b*Faculty of Medicine, University of Helsinki, Helsinki, Finland*

Abstract. The purpose of this study was to investigate the value mechanisms in implementing a digital health intervention (DHI) in different contexts and countries. We utilized realist evaluation and the CIMO logic (Context, Intervention, Mechanism, Outcome) to analyze the mechanisms explaining the value capture of Klinik Pro, an Intelligent Patient Flow Management system (IPFM), which is a DHI for seeking of treatment and triage purposes. The study was conducted as a multiple case study using semi-structured interviews to research four market expansions in three countries. In total, seven healthcare mechanisms were discovered: co-creation, proper competence level, coordination, evidence-based medicine, integration, proper timing, demand management. The first four mechanisms were the same in all cases. CIMO framework proved to be useful in the value formulation of the IPFM.

Keywords. Digital Health Intervention, Value-Based Healthcare, CIMO

1. Introduction

There are increasing investments into digital health interventions (DHIs), but the value capture of these solutions is difficult to measure due to the complexity of the interventions and their implementation areas [1-2]. In particular, the mechanisms through which the outcomes of the DHI are achieved are often elusive, because there is insufficient understanding of the dynamics between the implementation context and the activated value mechanisms in healthcare. Thus, this study focuses on expanding knowledge on how to discover the mechanisms that explain the value formulation of DHIs. The approach used for this is the CIMO logic, which stands for Context, Intervention, Mechanism, and Outcome [3]. As an approach emerging from design science, CIMO can be used to structure the logic of outcomes and dynamics of an intervention when implemented in a given context [4]. Value in healthcare in this study is understood as the health outcomes achieved per resources used, i.e. a relation [5].

The studied DHI in this research was an Artificial Intelligence based intervention by a Finnish health technology company, Klinik Healthcare Solutions Oy. Their DHI,

¹ Corresponding Author: Henni Tenhunen, Aalto University, Espoo, Finland; E-mail: henni.tenhunen@aalto.fi

called Klinik Pro, is an Intelligent Patient Flow Management system (IPFM) that supports efficient seeking of treatment and triage. Patients report their symptoms online to the IPFM, which then utilizes AI to analyze the level of urgency and suggesting an initial differential of diagnoses. This information is then directed to healthcare professionals to perform further actions (Figure 1).

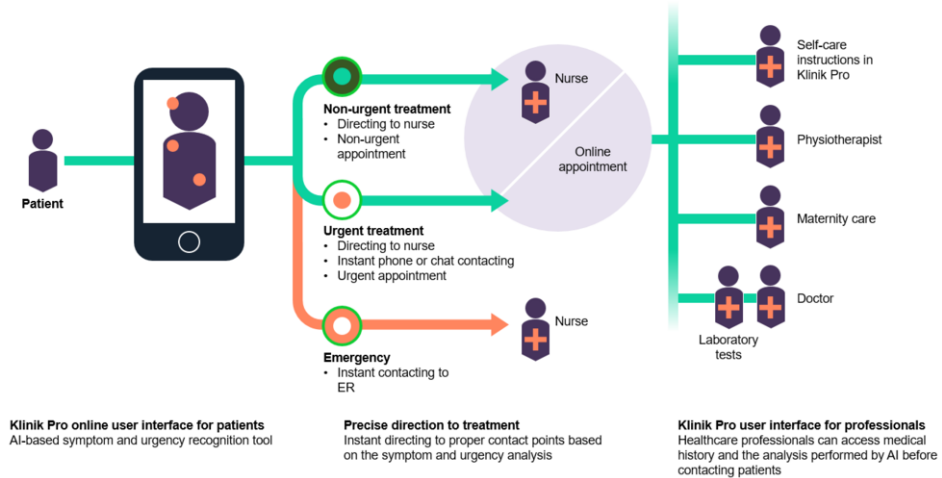


Figure 1. An example of a setup and functionalities of Klinik Pro (the IPFM)

The real-time empirical setting of this study was formed when in 2019 the company was simultaneously expanding its operations in four instances enabling international comparisons: one in Mexico, one in Portugal and two in Finland. These four contexts were analyzed using the CIMO logic and following [6]. By collecting data regarding the intervention, contextual environments, and expected outcomes of the IPFM, our aim was to deduce which value mechanisms are activated in the implementation of the IPFM. The main research question was: **RQ1: How to discover the mechanisms through which digital health intervention accomplishes value?** As a sub question we asked: **RQ1.1: How does the implementation context modify the activation of the mechanisms?**

2. Data and Methods

The research design followed the principles of design science and service engineering. Due to utilizing the CIMO logic in our empirical cases, the study also followed the principles of theory elaboration model [7] as well as grounded theory [8]. The study was performed as a multiple case study with one unit of analysis: the seeking of treatment and triage process within the primary care unit in each case (Table 1).

Table 1. Case descriptions.

Context	Targeted healthcare system	Main example of IPFM’s expected outcomes	Market phase
Large city in Finland	Public primary care	Reduce nurses’ stress levels	In Production
National association in Finland	Public primary care	More efficient handling of patient flows	Piloting / Pre-production
Large city in Mexico	Public and private primary care	Create a system to manage electronic health records	Pre-market

Medium large city in Portugal	Public primary care	Create a digital triage system	Piloting
-------------------------------	---------------------	--------------------------------	----------

We gathered data by conducting ten semi-structured interviews following the structure of the CIMO-configuration. The interviews took place during spring 2019 and were targeted on a diverse group of stakeholders within each market expansion including healthcare professionals, investors, managers, and IT experts (Table 2).

Table 2. List of informants.

Organization type	Description of the organization	Informant role
Health technology company	Klinik Healthcare Solutions	Chief Medical Officer
Primary care unit	Finnish medical center	Chief Physician
Primary care unit	Finnish medical center	Assistant Head Nurse
Primary care unit	Portuguese medical center	Managing Physician
Primary care unit	Portuguese medical center	Physician Intern
Primary care unit	Finnish medical center	Manager of Health Services
Primary care unit	Finnish medical center	ICT Service Coordinator
Venture capital	Healthcare VC company	CEO
Primary care unit	National healthcare association	CIO
Primary care unit	Mexican medical center	MD, CEO

3. Results

The activation of healthcare mechanisms relies not only on the dynamics of the intervention and its implementation, but also on the prerequisites of purposeful action. In all cases the informants considered the IPFM to improve the precision of information, break time and location boundaries, and enable inexpensive data collection from patients. However, there were differences in implementations as in Mexico the existing IT infrastructure could hinder proper use of the IPFM, whereas in Portugal the informants considered the decentralization of patient information to cause challenges. Considering the prerequisites for purposeful action, all informants considered the IPFM to provide improvements in the control information among healthcare professionals to make decisions that are more justified. Furthermore, the motivation to use the IPFM was perceived to be good in all cases. Finally, a few informants considered the IPFM to create new capabilities such as inventing new ways to plan clinical pathways utilizing the analysis performed by the IPFM’s AI.

Four healthcare mechanisms were discovered in all cases. First, *co-creation* is a mechanism that is inherently active in the IPFM due to engaging both patients and healthcare professionals in using the IPFM. Second, all informants perceived the IPFM to activate the *proper competence level* mechanism that allows the utilization of the lowest sufficient resource in each operation leading ultimately to cost savings. Third, the IPFM was considered to activate the *coordination* mechanism due to improved control information. Finally, as the IPFM’s AI performs the conditional analysis of the patients based on the most relevant scientific findings, the *evidence-based medicine* was the fourth mechanism active in all cases.

Besides the aforementioned four healthcare mechanisms, the Finnish informants considered the IPFM to support *integration* mechanism that results in more coherent and consistent knowledge of patients’ conditions. This was not considered active by the informants in the case of Mexico due to issues in the pre-existing IT infrastructure. Moreover, integration was also not considered to be activated in Portugal due to decentralized implementation. Besides integration, *proper timing*, a mechanism to

route and schedule patients effectively, was considered to be active in the Finnish cases. However, it was not seen activated in Mexico or Portugal due to same reasons as with integration. Finally, the seventh healthcare mechanism considered to be somewhat present in all cases was *demand management*. As the IPFM acts as an additional channel to seek treatment, all informants considered the possibility of the IPFM supporting efficient contacting of patients. However, in all cases it was also elaborated that this additional channel might result in overcapacity problems as patients might utilize multiple channels while seeking treatment. Thereby, the demand management mechanism was considered less relevant for all cases.

Generally, the mechanisms in each case were perceived to be very similar despite the contextual differences. In spite of the similarities, the varying intentions and priorities to utilize the IPFM among the stakeholders result in diverse outcomes responding to the actual problems the IPFM aims to solve. Thus, the same DHI can be used for very different purposes.

4. Discussion and Conclusions

By analyzing and comparing the cases, it is possible to discover patterns in the IPFM's functionality to understand its value formulation on a more general level. In short, the IPFM's ability to function correctly relies on the co-creation mechanisms to engage patients to report their symptoms to a digital system. Consequently, the practices regarding triage (e.g., whether it even exists) and the existing IT-infrastructure are the critical factors. Any issues with these factors result in uncertain activation of any healthcare mechanisms.

Depending on the willingness of actors, the IPFM produces a more precise understanding of the patient's condition, which forms a foundation for a better access, planning, and implementation of treatment. IPFM improves not only the knowledge of the patient and healthcare professionals but also enables optimization of resources, thus potentially leading to costs savings, enhanced productivity, and improved quality of care. Consequently, more effort can be allocated to actual patient care tasks. This is the kind of value that decision-makers appear to be looking for. Additionally, IPFM could even provide new competences for its users. However, a critical factor in realizing all the mechanisms rely on the users' adaptability and willingness to use the IPFM in the first place. Figure 2 illustrates this value formulation process of Klinik Pro (IPFM) on a general level. The purple color indicates patient while the red color refers to healthcare professionals. The IPFM and its activity is described with gray. Healthcare mechanisms are illustrated with blue while the outcomes are highlighted with green. The arrows indicate the interdependent relationships between the factors.

There are certain limitations in this research that need to be taken into account. First, this study concentrates merely on DHIs directed to triage and seeking of treatment, i.e. IPFM solutions. Second, the patients, the most important stakeholder group, were not studied, thus making patient perspective vital to study in future studies. Third, although being a multiple case study, the number of interviews in total was limited.

Despite these limitations, this study provides new insights into discovering the mechanisms of DHI value by examining the interdependent relationships between the intervention, its context, and its expected outcomes. Furthermore, this study supports the customers of the health technology companies (e.g. medical centers) to understand

the value formulation of similar DHIs as the IPFM, improving managerial decision-making. With this information, it is possible to concentrate on activities to maximize value capture while also detecting unnecessary activities.

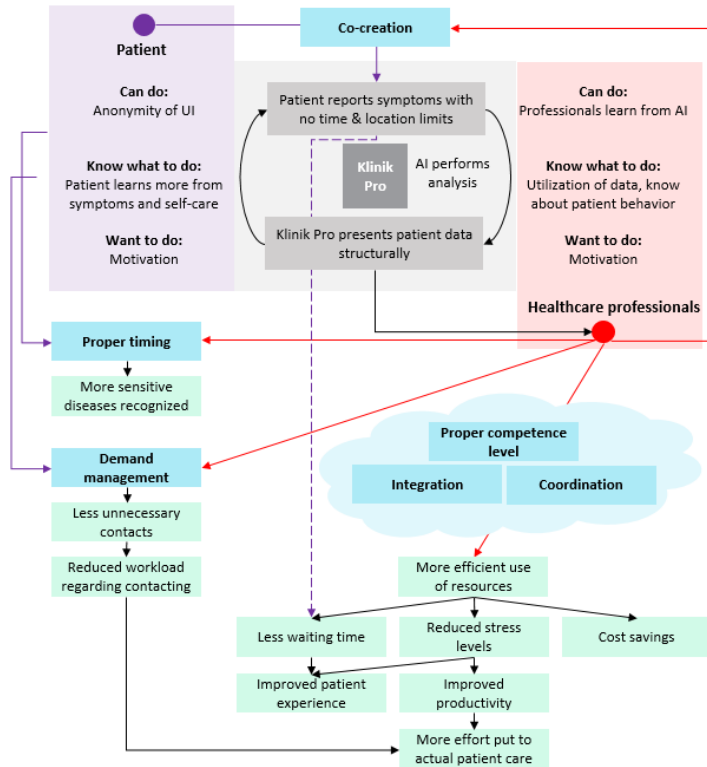


Figure 2. Value Formulation of Klinik Pro (the IPFM)

References

- [1] C. Sorenson, M. Drummond, B. Khan, Medical technology as a key driver of rising health expenditure: disentangling the relationship, *Clinicoecon Outcomes Res* **30** (2013), 223–234.
- [2] E. Murray, E. Hekler, G. Andersson, L. Collins, A. Doherty, C. Hollis et al., Evaluating Digital Health Interventions: Key Questions and Approaches, *American Journal of Preventive Medicine* **51** (2016), 843–851.
- [3] D. Denyer, D. Tranfield, J. Van Aken, Developing design propositions through research synthesis, *Organization Studies* **29** (2008), 393–413.
- [4] J. Holmström, T. Tuunanen, J. Kauremaa, Logic for design science research theory accumulation, *Proceedings of the Annual Hawaii International Conference on System Sciences* (2014), 3697–3706.
- [5] M. Porter, What is Value in Health Care?, *The New England Journal of Medicine* **363** (2010), 1–3.
- [6] A. Väljä, H. Tenhunen, H. Harno, L. Mäkitie, T. Tanila, P. Lillrank, From Tertiary to Primary Care – Understanding Context in the Transfer of Digital Headache Service Pathway, *Studies in Health Technology and Informatics* **262** (2019), 304–307.
- [7] M. Ketokivi, T. Choi, Renaissance of case research as a scientific method, *Journal of Operations Management* **32** (2014), 232–240.
- [8] B. Glaser, A. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine de Gruyter, New York, 1967.

Virtual Reality Therapy for Social Phobia: A Scoping Review

Elham SALEHI^{a,1} Mahdi MEHRABI^a Farhad FATEHI^{b,c} Asiyeh SALEHI^d

^a Islamic Azad University, Shiraz, Iran

^b Centre for Online Health, The University of Queensland, QLD, Australia

^c Tehran University of Medical Sciences, Iran

^d School of Health and Human Sciences, Southern Cross University, QLD, Australia

Abstract. Virtual reality (VR) has become an interesting alternative for the treatment of social phobia. This scoping review explored the impact of VR technologies in the treatment of social phobia (25 papers were included). Treatment approaches to social phobia included virtual reality exposure therapy, cognitive behaviour therapy, exposure group therapy, in vivo exposure therapy, and waiting list. VR was identified as the most successful approach, not only in improving social phobia, but also resulting in real life changes such as improvement in occupational success. VR Head Mounted Displays were used in most of studies (76%), followed by VR-based Specific Systems (24%). Although VR techniques indicated positive impact on patients to reduce their anxiety, it is paramount to consider variables such as the way therapists control the environment as well as individuals' characteristics for achieving the best outcome.

Keywords. Social phobia, Social fear, Social anxiety, Virtual reality

1. Introduction

Anxiety disorders are characterized by apprehension, tension, nervousness, or worry. Social phobia, as the most common anxiety disorder, is the unreasonable fear of social situations and the interaction with other people that can bring on feelings of self-consciousness, judgment, evaluation, and inferiority, and can result in social isolation [1]. Diagnosis and treatment of anxiety disorders are difficult as there are no objective biomarkers, however, phobia therapy is the first-line treatment for anxiety disorders, modifying the result of fear of in-memory structures [2, 3]. The treatment approaches include Cognitive Behavioural Therapy (CBT) and Exposure Therapy (ET). CBT includes exposure to fear-triggering stimuli (e. g. speaking in front of a group), social skills training and relaxation training to change cognition and behaviours [4]. In Exposure Group Therapy (EGT), patients are exposed to the feared object or context without any danger to overcome their fears. If the patient is exposed to the real situation, it is called In Vivo Exposure Therapy (iVET) [5].

Virtual Reality (VR) is a human-computer interaction that allows a person to step into an unseen environment to face their fears which participants could react in the virtual world with their observation through images on a computer screen [6, 7, 8, 9]. VR is an effective method for the treatment of a variety of phobias such as social phobia, post-

¹ Corresponding Author: Email: eliss.salehi@gmail.com

traumatic stress disorder, claustrophobia, arachnophobia and fear of heights [10, 11,12]. This review aimed to investigate the efficacy of VR in improving social phobia.

2. Methodology and design

This scoping review systematically searched four databases (PubMed, PsycINFO, IEEE Explore, and Science Direct) from 2000 to 2018. Keywords were: ("virtual reality" OR "virtual environment" OR "cyberspace") AND ("social phobia" OR "social isolation" OR "social fear" OR "anxiety"). Peer-reviewed papers written in English that used VR for improving any case of social phobia, social fear, and social speaking fear were included. A total number of 726 records were initially considered for the review, and a final set of 25 papers were deemed eligible and included in this scoping review. Constant comparative analysis was used to explore and categorize the themes across the studies. Synthesis of the information was performed based on different approaches of VR for treating social phobia, and comparing VR with conventional treatments.

3. Results

3.1 Study characteristics

Participants' age range was 18-70. Sample size varied from 2-97, with a relatively equal distribution of both genders. Treatment included Virtual Reality Exposure Therapy (VRET) with the highest percentage (65%), cognitive behaviour therapy (CBT), exposure group therapy (EGT), in-vivo exposure therapy (iVET) and waiting list (WL). This review identified three main themes.

3.2 Virtual reality in the treatment of social phobia - trends and applications

By comparing different therapeutic approaches, VR approaches and CBT showed the most remarkable improvements in the treatment of social phobia [1, 5-6, 13-18]. There were no significant differences between VR and CBT and they indicated relatively equal outcomes [1, 6]. However, VR was highly recommended and there has been an increasing number of studies during the last 19 years. VR is affordable and provides equal or even more satisfactory results in comparison to other non-VR approaches such as CBT. VR has a paramount impact on not only the treatment of social phobia but also different aspects of personal life and wellbeing. Some examples are: more improvement in well-being and quality of life; fewer sessions for accomplishing the therapy; confidence [14], positive self-statement [5], improvement in emotion and communication, real life changes such as occupational success [7, 17], decreased fear, anxiety, avoidance of the social situation, improvement in physiological measures on public speaking anxiety, patient's attention focus, and changing behaviour [10, 14, 15].

The number of treatment sessions in VR interventions varied from 1-12 sessions and the duration ranged from 5-90 minutes. There were also variations in the follow-up period, from a few weeks to one year. Among VR approaches, iVET group was more realistic than VR conversation, and participants commented that the VR conversation would be more realistic if they did not meet the actor and if they were not in the same

room as the other participants [1, 6, 16, 19]. In addition, fear rating was higher during VR conversation than in iVET conversation [1, 6, 16, 19].

3.3. Virtual reality tools for social phobia treatment

The VR equipment can be categorized into two groups: Virtual Reality Head Mounted Displays (HMD), and Virtual Reality based Specific Systems (VRBSS). The majority of the research experiments (76%) used HMD, while, the other 24% of the studies used VRBSS, mainly different 3D display/sound equipment and motion trackers.

For HMD usage, there is a Distributed Interactive Virtual Environment, which consists of a computer, HMD, display screens for each eye, stereo ear-phones, Silicon Graphics, main memory, avatar face and model and a head-tracking system. In a virtual environment (VE), the conversation is managed through a virtual avatar interaction with participants, the virtual social situations such as shopping; attending a job interview include semi-scripted dialogues consists of modification templates based around the video game. The participants interact via HMD and wear headphones and a microphone headset to interact. Therefore, avatars mouth movements during the dialogue seem natural [19, 22, 23]. The tracker allows patients to see the virtual world with head movement in the real world, and software package that provides the scenes of speech among the audience [5, 16]. The patient's head movements change the environment, and the therapist can control the audience's reactions and communicate with them via a microphone. During the talk, their heart rate is monitored and the content of the presentation is recorded [10, 12, 14, 15, 20, 21]. Due to the limitation of VR hardware and software, and the associated expenses, off-the-shelf applications for different issues, such as public speaking practice during different scenarios have been developed and are available on mobile app stores [24, 25, 26].

VRBSS consists of the Delft Remote VR system with the capacity to display 3D video, 3D images and 3D glasses and ear headphones for sound immersion. The program displays images and then commands given by the therapist through a wireless USB keyboard for activation of the interaction and changing the scene [17, 23]. The laboratory for VRBSS comprises of two rooms that are separated by a mirror while the therapist can see the patient but the patient cannot see the therapist, they have a face-to-face contact during exposure and communicate via intercom [27].

3.4 Measurement tools in social phobia

Two types of measurements were used during the course of treatment: 1) self-report measure - individual's own report of their symptoms, behaviours, beliefs, or attitudes. This was conducted via pre-treatment and post-treatment assessments. 2) physiological measure - measuring bodily variation such as: heart rate, dry mouth, and blood pressure. An overall of 30 psychological instruments were used as outcome scales across studies [11]: APQ (answer psychology questions), ATQ30 (automatic thought questionnaire), ATPS (attitudes towards public speaking questionnaire), BDI (Beck depression inventory), BCSS (brief core schema scales), BAT (behavioural assessment task), CGI (clinical global improvement scale), DAS (dysfunctional attitude scale), DASS (depression anxiety stress scale), DSM-IV (diagnostic and statistical manual of mental disorder), EUROHISQOL (Eurohis quality of life scale), ERQ (emotion regulation questionnaire), FNE-B (fear of negative evaluation-brief Form), GPTS (green paranoid thought scales), HAD (hospital anxiety and depression), HADS (hospital anxiety and

depression scale), IPQ (I group presence questionnaire), LSAS (Liebowitz social anxiety scale), LSAS-SR (Liebowitz social anxiety scale-self report), PRCA (personal report of communication apprehension), PDBQ (personality disorder belief questionnaire), PRCS (personal report of confidence of speaker), PDBQ (personality disorder belief questionnaire), SUS (slater-Uschsted questionnaire), SCID (structured clinical interview for the DSM-IV), SSPS (positive and negative self-statement during public speaking), SCIA (social contexts inducing anxiety), SIAS (social interaction anxiety scale), SAS (social adjustment scale), SF -36 (the medical outcome 36-item short-form health survey), SSQ (simulator sickness questionnaire), SPAI (social phobia and anxiety inventory), STAI (state-trait anxiety inventory), and SUD (subjective unit of discomfort) [1-30]. Key methodological features and findings of the reviewed studies are available online ([DOI: 10.13140/RG.2.2.12587.39209/1](https://doi.org/10.13140/RG.2.2.12587.39209/1)).

4. Discussion and recommendations

VR treatment has a great impact on the communication and seems to be more effective and efficient than conventional therapies for social phobia. However, there have not been any specific comparison between different VR tools, and few mobile-base tools were considered. In addition, from a technological aspect, although HMD and VRBSS are the key effective VR tools, there has not been any specific comparison between them to examine the outcome and performance clearly.

Although the diverse variety of VR tools indicated some type of positive impact in reducing the anxiety, it is still paramount to consider confounding variables, such as the individuals' characteristics and the way therapists control the environment. Some of the participants felt anxious during VRET, due to the presence and the monitoring of the therapist during the session. In addition, therapists should try to maximize a patient's presence because attentional focus leads to better treatment. A powerful sense of presence and immersion is needed to make the experience real, and to enhance the effectiveness of the interventions [6, 21]. Further studies (with larger sample size) are required to clarify the differences between different VR tools in regards to their effectiveness and efficiency. There have been also limited studies around mobile-based VR and their potentials [24, 29, 30]. Future studies may also explore different behavioural patterns in regards to gender for using VR and the treatment outcomes.

There were issues related to the interventions that need to be considered to facilitate future studies. These include: difficulties in producing and editing movies as they are time-consuming [30]; different behavioural patterns due to gender; lack of control for audience interactions with each other for the virtual reality-based specific system; lack of sense of presence/immersion to overcome cognitive barriers [29]; and patients' anxiety due to therapist presence/monitoring [22], small sample size, and limiting the generalizability of the results of the studies [14].

References

- [1] Klinger, E., et al., Virtual reality therapy versus cognitive behavior therapy for social phobia: a preliminary controlled study. *Cyberpsychol Behav*, 2005. 8(1): p. 76-88.
- [2] Rose, M. and J. Devine, Assessment of patient-reported symptoms of anxiety. *Dialogues in clinical neuroscience*, 2014. 16(2): p. 197.

- [3] Khosravi, P., A. Rezvani, and A. Wiewiora, The impact of technology on older adults' social isolation. *Computers in Human Behavior*, 2016. 63: p. 594-603.
- [4] Klinger, E., Chemin, I., Légeron, P., et al., Definition of a VR-based protocol to treat social phobia. *CyberPsychology & Behavior* 6:411-420., 2003.
- [5] Safir, M.P., H.S. Wallach, and M. Bar-Zvi, Virtual reality cognitive-behavior therapy for public speaking anxiety: one-year follow-up. *Behav Modif*, 2012. 36(2): p. 235-46.
- [6] Anderson, P.L., et al., Virtual reality exposure therapy for a social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 2013. 81(5): p. 751-760.
- [7] Bush, J., Viability of virtual reality exposure therapy as a treatment alternative. *Computers in Human Behavior*, May 2008. 24(3): p. 1032-1040.
- [8] Roy, S., et al., Definition of a VR-based protocol to treat social phobia. *Cyberpsychol Behav*, 2003. 6(4): p. 411-20.
- [9] Klinger, E., et al., Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study. *CyberPsychology & Behavior*, 2005. 8(1): p. 76-88.
- [10] Anderson, P.L., et al., Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depress Anxiety*, 2005. 22(3): p. 156-8.
- [11] Alsina-Jurnet, I., The role of presence in the level of anxiety experienced in clinical virtual environments. *Computers in Human Behavior*, January 2011. 27(1): p. 504-512.
- [12] Cornwell, B.R., et al., Anticipation of public speaking in virtual reality reveals a relationship between trait social anxiety and startle reactivity. *Biol Psychiatry*, 2006. 59(7): p. 664-6.
- [13] Lister, H.A., C.D. Piercey, and C. Joordens, The effectiveness of 3-D video virtual reality for the treatment of fear of public speaking. *Journal of CyberTherapy and Rehabilitation*, 2010. 3(4): p. 375-381.
- [14] Anderson, P., B.O. Rothbaum, and L.F. Hodges, Virtual reality exposure in the treatment of social anxiety. *Cognitive and Behavioral Practice*, 2003. 10(3): p. 240-247.
- [15] Harris, S.R., R.L. Kemmerling, and M.M. North, Brief virtual reality therapy for public speaking anxiety. *Cyberpsychol Behav*, 2002. 5(6): p. 543-50.
- [16] Wallach, H.S., M.P. Safir, and M. Bar-Zvi, Virtual reality cognitive behavior therapy for public speaking anxiety: A randomized clinical trial. *Behavior modification*, 2009. 33(3): p. 314-338.
- [17] Gebara, C.M., et al., Virtual reality exposure using three-dimensional images for the treatment of social phobia. *Rev Bras Psiquiatr*, 2016. 38(1): p. 24-9.
- [18] Kampmann, I.L., et al., Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial. *Behav Res Ther*, 2016. 77: p. 147-56.
- [19] Powers, M.B., et al., Do conversations with virtual avatars increase feelings of social anxiety? *Journal of anxiety disorders*, 2013. 27(4): p. 398-403.
- [20] Slater, M., et al., An experimental study on fear of public speaking using a virtual environment. *CyberPsychology & Behavior*, 2006. 9(5): p. 627-633.
- [21] Price, M., et al., Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. *J Anxiety Disord*, 2011. 25(6): p. 763-70.
- [22] Morina, N., et al., Sense of presence and anxiety during virtual social interactions between human and virtual humans. *PeerJ*, 2014. 2: p. e337.
- [23] Brinkman, W.-P., et al., A virtual reality dialogue system for the treatment of social phobia. 2012: ACM.
- [24] Lindner, P., et al., Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: A randomized controlled trial. *Journal of anxiety disorders*, 2018.
- [25] Stupar-Rutenfrans, S., L.E.H. Ketelaars, and M.S. van Gisbergen, Beat the Fear of Public Speaking: Mobile 360° Video Virtual Reality Exposure Training in Home Environment Reduces Public Speaking Anxiety. *Cyberpsychology, Behavior, and Social Networking*, 2017. 20(10): p. 624-633.
- [26] Kim, H.E., et al., Effectiveness of self-training using the mobile-based virtual reality program in patients with a social anxiety disorder. *Computers in Human Behavior*, 2017. 73: p. 614-619.
- [27] Kampmann, I.L., et al., Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial. *Behaviour Research and Therapy*, 2016. 77: p. 147-156.
- [28] Anderson, P.L., S.M. Edwards, and J.R. Goodnight, Virtual Reality and Exposure Group Therapy for Social Anxiety Disorder: Results from a 4-6 Year Follow-Up. *Cognitive Therapy and Research*, 2017. 41(2): p. 230-236.
- [29] Gega, L., et al., Virtual environments using video capture for social phobia with psychosis. *Cyberpsychol Behav Soc Netw*, 2013. 16(6): p. 473-9.
- [30] Jo, H.J., et al. Movie-based VR therapy system for treatment of anthropophobia. in *Engineering in Medicine and Biology Society*, 2001. Proceedings of the 23rd Annual International Conference of the IEEE. 2001.

Visualizing the Cascade Effect of Redesigning Features in an EMR System

Samar HELOU^{a,1}, Victoria ABOU-KHALIL^b, Goshiro YAMAMOTO^c,
Osamu SUGIYAMA^d and Tomohiro KURODA^{b,c}

^aGlobal Center for Advanced Medical Engineering and Informatics, Osaka University

^bDepartment of Social Informatics, Graduate School of Informatics, Kyoto University

^cDivision of Medical IT & Administration Planning, Kyoto University Hospital

^dPreemptive Medicine and Lifestyle Related Diseases Research Center, Kyoto University

Abstract. Electronic Medical Record (EMR) systems are complex systems with interdependent features. Redesigning one feature of the system can create a cascade effect affecting the other features. By calculating the cascade effect, the designers can understand how each individual feature could be affected. This understanding allows them to maximize the positive effects and avoid negative consequences of their redesign activities. To understand the cascade effect, the designers can look at their computations' results; a task that becomes more difficult when the number of features grows. To reduce their task load, we propose a tool for visualizing the cascade effect of redesigning features in an EMR system. Our preliminary evaluation with six graduate students shows that visualizing the cascade effect reduces the task load and slightly improves their performance when analyzing the cascade effect. Ways for improving the tool include (i) showing the computation results within the visualization, and (ii) allowing the designers to compare the cascade effect generated by redesigning different features.

Keywords. Design, EMR System, Visualization

1. Introduction

The designers of Electronic Medical Record (EMR) systems need to continuously redesign the systems to fulfill the emerging needs of their users [1,2]. It is common for the designers to be faced with multiple EMR features to redesign. Each EMR feature represents a distinctive aspect or attribute of the EMR system. For example, one EMR feature could be its usability, while another EMR feature could be its integration with the used medical devices. Since the designers have limited resources, they need to choose which feature to redesign. This choice is difficult because EMR systems are complex socio-technical systems where the systems' features are interdependent i.e., redesigning one feature will affect the other features of the system. For example, integrating the EMR system with the used medical devices would increase its usability. As another example, automatically generating EMR summaries from the EMR notes would facilitate the exchange of EMRs between healthcare providers which would consequently reduce the preparation time for the consultations [3].

¹ Corresponding Author: Samar Helou, E-mail: helou.samar@gmail.com.

Previous work showed that considering the interdependency of the features is needed to know which features are the most important to (re)design [3]. By calculating the cascade effect generated by (re)designing each feature, the designers can assess the overall effect that redesigning each feature will have on the system as a whole. Accordingly, the designers can optimize their design strategies by choosing to work on the features that generate the largest positive effect on the system.

To understand the cascade effect of redesigning each feature, the designers need to understand how each individual feature in the system is affected. To do so, the designers can check the details of the computations. However, when multiple features are considered, looking at the computations' outputs requires a high mental demand and could result in an incomplete understanding. In this case, visualizing the cascade effect can reduce the required mental workload by offloading work to the perceptual system [4].

The aim of this work is to support EMR designers in understanding the cascade effect of their redesign activities. To this end, we propose a tool for visualizing the cascade effect generated by redesigning features in an EMR system.

2. Methods

Previous work described a method for calculating the cascade effect of redesigning features in an EMR system [3]. First, the designers have to analyze the relationships between the features and then compute the overall effect of modifying each feature.

2.1. Analyzing the Relationships between the EMR Features

First, the designers have to analyze and quantify the influence that each feature has on all the other features. For every two features f_i and f_j , the designers have to answer the following question: "If we redesign feature f_i , how will that influence feature f_j ?"

To answer this question, the designers assume that feature f_i has a positive influence on feature f_j when an improvement of f_i leads to an improvement of f_j . Similarly, feature f_i has a negative influence on feature f_j when an improvement of f_i leads to a deterioration of f_j . To quantify the influence, the designers can employ a linear scale ranging from "-0.9: very strong negative influence" to "0.9: very strong positive influence" with the middle point being "0: no influence".

Once the relationships between the features are quantified, the designers can create the feature dependency matrix $D = [d_{ij}]$ where all principal diagonal elements are equal to zero and d_{ij} represents the degree to which feature f_i has an influence on feature f_j .

2.2. Computing the Cascade Effect

$A = [a_{ij}]$ is the direct effect matrix where a_{ij} represents the direct effect of modifying feature f_i on feature f_j . Matrix A is an identity matrix. The cascade effect of modifying each EMR feature is computed using the following equation:

$$C = A \times D + A \times D^2 + \dots + A \times D^h, h \rightarrow \infty$$

$C = [c_{ij}]$ where c_{ij} is the cascade effect that modifying feature f_i has on feature f_j .

2.3. Visualizing the Cascade Effect

The purpose of the visualization tool was to support the designers in understanding the effects of redesigning a feature. Therefore, for every feature, we needed to visualize the cascade effect that redesigning this feature will induce on all the other individual features. To build a prototype of the visualization tool, we worked with a case study of redesigning an antenatal care EMR system. We used the list of features and the dependency map presented in that study [3]. To compute the cascade effect of redesigning each feature, we used python and followed the calculation steps described in section 2.2.

We built the prototype of the visualization tool using HTML, CSS, and JavaScript. The prototype’s interface is shown in Figure 1. The green bars represent the baseline, i.e., the initial state of the EMR system before any redesign activity. By pressing on a “Redesign” button, the users can see the direct effect in blue and the indirect effects in yellow. The figure shows, as an example, the cascade effect of redesigning the EMR system to automatically generate summaries from the EMR notes. The working prototype is accessible via: casceffect.netlify.com.

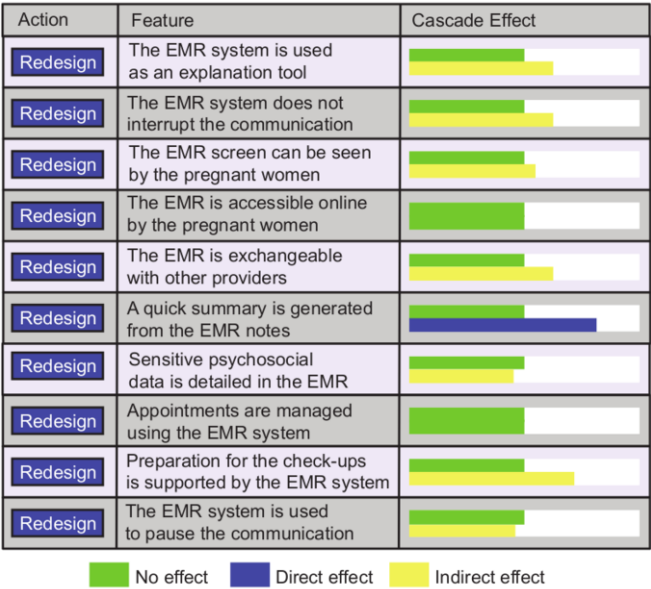


Figure 1. The interface of the visualization tool.

2.4. Evaluating the Prototype of the Visualization Tool

We conducted a preliminary evaluation of the visualization tool with six graduate students. We indicated to the participants the three features that have the largest overall effect on the system. Then, we asked them to conclude why these three features had the largest overall effect by using the results of the computations shown in Figure 2, and by using the visualization tool. Half the participants used the results of the computations first and the other half used the visualization tool first. After using each method, the participants answered a NASA-TLX questionnaire. NASA-TLX is a multi-dimensional scale used to measure the workload estimates of a task [5].

The EMR system is used as an explanation tool	1.00	0.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.00
The EMR system does not interrupt the communication	0.00	1.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.00
The EMR screen can be seen by the pregnant women	0.00	0.00	1.11	0.22	0.11	0.03	-0.37	0.00	0.00	0.00
The EMR is accessible online by the pregnant women	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
The EMR is exchangeable with other providers	0.00	0.00	0.11	0.22	1.11	0.33	-0.37	0.00	0.00	0.00
A quick summary is generated from the EMR notes	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Sensitive psychosocial data is detailed in the EMR	0.00	0.00	-0.37	-0.73	-0.37	-0.11	1.22	0.00	0.00	0.00
Appointments are managed using the EMR system	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Preparation for the check-ups is supported by the EMR system	0.00	0.00	0.03	0.07	0.33	0.70	-0.11	0.00	1.00	0.00
The EMR system is used to pause the communication	0.00	-0.30	-0.33	-0.07	-0.03	-0.10	0.11	0.00	0.00	1.00

Figure 2. The computations results.

3. Results

The results of the evaluation are shown in Figure 3. Each bar represents the distribution of the scores given by the six participants. By considering the mean and the median of the reported scores, we conclude that using the visualization tool tends to require less mental demand, temporal demand, and effort. Moreover, using the visualization tool may reduce the designers’ frustration and slightly increase their perceived performance. However, the tool requires more physical demand due to its interactive nature.

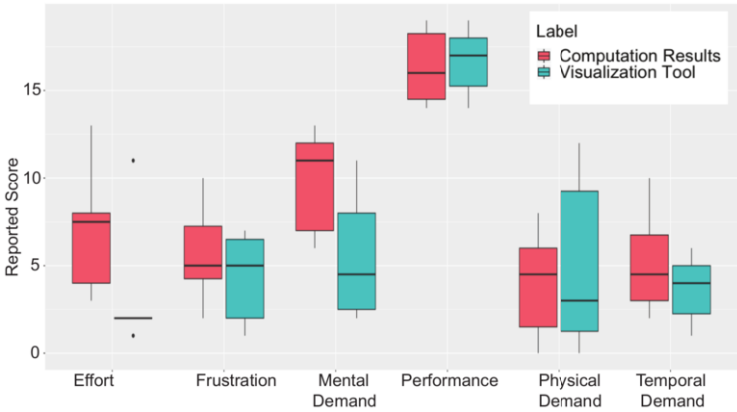


Figure 3. Using the computations’ results versus using the visualization tool.

4. Discussion

All the participants agreed that using the visualization tool required less mental and temporal demand than the computations’ results. However, two participants reported that the visualization tool was more physically demanding. One of these two participants further reported that using the tool required more effort and caused them more frustration. The effort required by the tool was attributed to the difficulty of

comparing the cascade effect of different features. The current prototype only allows the designers to visualize the cascade effect of one feature at a time. One way to address this issue is to allow the designers to choose multiple features and visually compare their cascade effects.

The participants also reported that they could not accurately quantify the cascade effect when they used the tool. Instead, they had to qualitatively and approximately assess it. This issue could be addressed by adding the quantitative data from the computations' results to the visualization. Moreover, in order to fully understand the cascade effect, it is important to know how the effect is propagating across the features' network. In other terms, the designers need to know the source of the effect on each feature. To this end, an animation tool showing how the cascade effect propagates along the system's features might prove fitter as these tools are usually used to represent dynamic systems [6].

Finally, it is important to note that the effect generated by a design change depends on the way the features are implemented and on the context in which the EMR system is used. Moreover, the calculated cascade effect depends on the subjective interpretations of the designers. Therefore, the prediction of the cascade effect may not accurately reflect the real world. The method's current merit lies in its ability to support the designers in finding a way of calculating, approximately, where a good course of redesign lies. Further research is needed to evaluate the method in work place settings.

5. Conclusion

Understanding the cascade effect of redesigning features in an EMR system allows designers to optimize their redesign strategies. To support the designers in their analysis, we proposed a tool to visualize the cascade effect of redesigning features of an EMR system. Our preliminary evaluation with six participants showed that the visualization tool could reduce the designers' mental and temporal demand and improve their performance when analyzing the cascade effect of redesigning the system's features. Further work is needed to reduce the physical demand required by the visualization tool and to allow designers to easily compare the cascade effect of redesigning different features.

References

- [1] M. Jaspers, L. Peute, A. Lauteslager, P. Bakker, Pre-post evaluation of physicians' satisfaction with a redesigned electronic medical record system, *Stud Health Technol Inform* **136** (2008), 303-308.
- [2] S. Helou, V. Abou-Khalil, G. Yamamoto, E. Kondoh, H. Tamura, S. Hiragi, O. Sugiyama, K. Okamoto, M. Nambu, T. Kuroda, Understanding the situated roles of electronic medical record systems to enable redesign: Mixed methods study, *JMIR Human Factors* **6** (2019), 13812.
- [3] S. Helou, V. Abou-Khalil, G. Yamamoto, E. Kondoh, H. Tamura, S. Hiragi, O. Sugiyama, K. Okamoto, M. Nambu, T. Kuroda, Prioritizing features to redesign in an EMR System, *Stud Health Technol Inform* **264** (2019), 1213-1217.
- [4] M. Tory, T. Moller, Human factors in visualization research, *IEEE Transactions on Visualization and Computer Graphics* **10** (2004), 72-84.
- [5] S. Hart, NASA-task load index (NASA-TLX); 20 years later, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **50** (2006), 904-908.
- [6] M. Hegarty, S. Kriz, C. Cate C, The roles of mental animations and external animations in understanding mechanical systems, *Cognition and Instruction* **21** (2003), 209-249.

Web Application for Home Care Providers

Virginia SANDULESCU^{a,1}, Sorin PUSCOCI^a, Monica PETRE^a, Alexandru GIRLEA^b
and Viorel BOTA^c

^a*Department of Telematics and Remote Control, INSCC, Bucharest, Romania*

^b*QuickWeb Info SRL, Bucharest, Romania*

^c*Elisa Med, Alba Iulia, Romania*

Abstract. The paper presents a web application integrated in a platform containing hardware and software components for optimizing the activities of home care providers (HCPs) and their services. The integrated platform is called CDMS, an acronym for the Romanian translation of Dispatching and Management Center for optimizing the Services offered by HCPs. The presented web application is at the center of the platform, offering functionalities for optimizing the activities of an HCP, the activities of the nursing staff and other functionalities that will be presented in the paper.

Keywords. home care provider, web application for home care, home nursing

1. Introduction

Population aging can be defined as an increase in the proportion of older persons that changes the age composition of a population [1] and it is already well known that there is an ongoing process of aging for the population of the entire globe. Europe has the oldest population (21% of its population aged over 60).

Aging usually implies some level of decrease in physical and mental capacity and a growing risk of disease [3]. Common health problems for elders include mobility problems associated with pain and frailty, hearing/ visual acuity reduction, diabetes, depression, Alzheimer, dementia [3]. These implications are the reason that most elders benefit from continuous/ periodic care: health care and assistance in performing certain tasks.

The best motivation in developing solutions for health and social care for elders arises from studying the need (the aging population) and is driven by the altruist purpose of taking care of our elders: our older relatives. Other motivation is the fact that healthy and more independent elders can lower the burden felt on public and private health sectors and they can also contribute to the economy and the society that integrates them (by working as mentors, caregivers, innovators, entrepreneurs or other members of the workforce [3]). The less altruist reason to develop solutions for elders is the fact that a market dynamics analysis shows that older adults are projected to spend \$15 trillion annually by 2020 and they will generate more than a third of global consumption growth [4]. [4] recognizes the caregiving market as a market that offers great opportunities, as it is expected to reach \$72 billion by 2020.

¹ Corresponding Author, Virginia Sandulescu, Department of Telematics and Remote Control, INSCC Bucharest, Preciziei 6, Bucharest, Romania; E-mail: virginia.sandulescu@inscc.ro.

The paper presents a web application integrated in a platform containing hardware and software components for optimizing the activities of home care providers (HCPs) and their services. The integrated platform is called CDMS, an acronym for the Romanian translation of Dispatching and Management Center for optimizing the Services offered by HCPs. The CDMS platform is an integrated platform containing software components: web applications acting as access portals for different types of users (a web application for users at the CDMS level and a web application for the users of HCPs: employees with different roles and permissions and patients/ informal caregivers), scripting applications for technical system administration, mobile apps for point of care documentation and supply management, APIs for integrating teleassistance/ telemonitoring devices and hardware components that allow running all the above mentioned software components. The designed functionalities for the software components are presented in [5]. The current paper focuses on the web application developed for the HCPs. The presented web application is at the center of the platform, offering functionalities for optimizing the activities of an HCP, the activities of the nursing staff and other functionalities that will be presented in the following sections. It is designed to be accessed as a Platform as a Service, so that an HCP subscribes to a CDMS and gets access to all the functionalities it needs to run.

2. Technologies used for development

Container technologies have been used to allow for a clear delimitation between data and processing and for a clear delimitation between the components dedicated to each subscribed HCPAs a security measure, the database and the web application are deployed on separate containers, to allow for better access control on the data. Another reason for using containers is for obtaining separate modules that may easily be replicated. This way, accommodating a new HCP consists, mainly, in replicating two containers: the data and the processing containers. An in-house container orchestration solution has been developed to allow the management of containers and the necessary services for each of them.

The chosen main programming language is PHP, a well-known and used programming language for web portals, for managing dynamic content, managing database connections, keeping track of opened sessions and so on. Laravel [6] is an open source development framework for PHP and it was chosen for this project. It uses the Model-View-Controller design pattern [7], it allows modular development, easy code reutilization, features that aid the testing activities, routing between the components of the web application, configuration management, template utilization, implementation of predefined security features and many other features. The database is created using MySQL, one of the most popular open source database management systems.

3. Implemented features

The web application for HCPs has been developed keeping in mind the need for integration in the whole CDMS platform.

The first step in developing the web platform consisted in analysis of the existing framework of HCPs. This stage contained documentation based on available literature

(legislation, normatives, case-studies) but also on field visits, interviews and open discussions with employees of HCPs. The specification design document summarized requests directly from the end-users.

The web application offers the following main features: management of the HCP's presentation web page; management of the HCP's service offer; employees management; clients management; inventory management; management of medical data; management of nursing visits; management of authorized devices: mobile devices for point-of-care documentation and devices for teleassistance/ telemonitoring. The web application contains predefined lists of procedures and services together with default supplies needed for each of the physiological parameters to be monitored, but it also allows for customization of these lists by allowing the input of new types of procedures and alterations on their characteristic data. A short description of the designed features is available in [5]. Table 1 presents most of the functionalities offered through the main navigation menu of the web application.

Table 1. Navigation menu (partial) of the web application for HCPs

Company configuration	
HCP profile	general information: legal name, address, logo, main contact information and other info
presentation web site	for each HCP a presentation website is generated in a predefined format: main page/ about page/ contact page/ terms and conditions for using the application
bank accounts	configuration of bank accounts used for issuing bills and other financial aspects
working addresses	configuration of working addresses
Employees' management	
manage employees list	create, view, edit employees' accounts; activate/ deactivate accounts
Patients management	
manage patients list and their electronic health records	view/ add/ edit/ deactivate patient account for each of the patients, manage: - personal data, medical record, physiological measurements, uploaded files, medical prescriptions, requested services (subscriptions), appointments
Devices management	
manage list of authorized devices	view/ add/ edit/ remove authorized devices that may access the platform (mobile devices for point of care documentation and telemonitoring/ teleassistance devices)
Supplies management	
manage list of supplies	view/ add/ edit/ remove supply item view/ add/ edit/ remove supplier view/ add/ edit/ remove producer of supplies manage current inventory inventory logs
Medical services management	
manage list of offered medical services	view/ add/ edit/ remove offered procedures view/ add/ edit/ remove managed health problems view/ add/ edit/ remove allergy types view/ add/ edit/ remove physiological parameters view/ add/ edit/ remove type of documents that may be added to a patient's electronic health record (medical report, test results, treatment plans, consent forms and others, medical prescriptions)
Subscriptions management	
manage subscriptions and offers	manage subscription types and packages of services manage offered services (other than medical)
Financial management	
manage financial documents	manage billing information/ invoices/ pricing
Calendar management	

schedule management	appointments management employees working schedule
Reports management	
generate reports	employee's activity patient's services supplies reports performance reports financial reports
Notifications management	
manage notifications for:	unpaid bills appointments: requested/ missed appointments support notifications: tickets generated by employees or patients vital parameters notifications: parameters with abnormal values supplies alerts: out of stock items or items below a predefined threshold subscriptions: expired or expiring subscriptions

These features are designed to be accessed by the following user categories: a) administrative personnel; b) nursing staff; c) clients/ patients of an HCP and their informal caregivers. These three large categories of users are identified in the activity of any HCP. In order to create an efficient (secure) access mechanism to the resources and functionalities of the web application, a role and permission based access control list has been implemented. The following roles/ activities have been identified and implemented in the platform: IT/ technical administrator; general administrator; financial manager; human resources manager; medical manager; medic; nurse; telemonitoring/ teleassistance manager. As observed from the field activities performed during the specification design phase and continued during development, different HCPs have different organizational charts, so these roles may be differently implemented in each HCP: for example an HCP may only have a small number of administrative employees, so the same person may perform multiple roles, or a larger HCP may have multiple employees for financial management so each of them must perform only a subset of the tasks involved in this specific activity. In order to accommodate that, a more specific list of permissions has been created, mainly a permission for each of the main functionality of the platform (management of the presentation web site, of the bank accounts, employees, billing and taxes, supplies, general medical data management, patients, appointments and so on). With this level of customization, a user's access to the application's data and functionalities may be accurately controlled.

4. Results

The web application has been deployed together with the other components of the CDMS platform at a center for managing HCPs in Alba County, Romania. The center is acting both as a CDMS and as an HCP. As seen in practice, the administrative personnel reported that the platform is easy-to-use, has an intuitive interface and this has been shown through the fact that it was a matter of a few months for the administrative personnel to integrate the usage of the platform in their activities. At the moment, the experimental deployment of the web application is used for keeping track of the inventory of the HCP, the employees list and test patients. Testing activities are still currently being run for each of the functionalities presented in the paper leading to functionalities improvements and minor bug fixes.

5. Conclusions and future work

The paper presents a web application for optimizing the activity of HCPs and the services they offer. The relevance of the proposed solution is increased in the context of an aging population with complex care needs that would benefit from home care services. Also, the market created for elders is expected to be the fastest growing market in the following years.

The application represents a single point of access for managing all the administrative activities of an HCP, including managing the nursing activities. There is no installation required from the point of view of an HCP. The CDMS offers the web application and the complex hardware requirements for storage and processing (in the situation of a complex HCP). The HCP only needs any kind of Internet enabled terminal to access the web application. The most important difference between the proposed solution and readily available solutions on the market, is the integration in the whole CDMS platform that offers additional components and functionalities, the most important of them include a mobile application for point-of-care documentation, a mobile app for supplies management and software components for technical management. Another important aspect of the proposed solution is the integration of an API that allows the platform to gather health data from devices provided by different vendors already on the market.

Acknowledgment

This work has been supported through the Project: Dispatching and Management Center for Optimizing Home Care Integrated Services, from Operational Competitiveness Program 2014-2020 Axis 1 – Romanian Research Ministry.

References

- [1] K. Land, V. Lamb, *International Encyclopedia of Public Health (Second Edition)*, Elsevier, 2017
- [2] R. Atoyan, L. Christiansen, A. Dizioli et al., *Emigration and Its Economic Impact on Eastern Europe*, International Monetary Fund, 2016
- [3] World Health Organization, *The Global Strategy and Action Plan on Ageing and Health*, 2017
- [4] M. Arensberg, Population aging: opportunity for business expansion, an invitational paper presented at the Asia-Pacific Economic Cooperation (APEC) International Workshop on Adaptation to Population Aging Issues, July 17, 2017, Ha Noi, Viet Nam, *J Health Popul Nutr* **37** (2018), 7.
- [5] V. Sandulescu, S. Puscoci, M. Petre, S. Soviany, M. Chirvasa, A. Girlea, Dispatching and management center for optimizing home care integrated services, in M. Ziefle & L. Maciaszek (eds.), *Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health* (2019), 255-261
- [6] Laravel, The PHP Framework for Web Artisans, available at <https://laravel.com/>.
- [7] MVC – MDN Web Docs Glossary: Definitions of Web-related terms, available at <https://developer.mozilla.org/en-US/docs/Glossary/MVC>

This page intentionally left blank

Section 3

Health and Prevention

This page intentionally left blank

A Model for Early Identification of Chronic Obstructive Pulmonary Disease

Thomas KRONBORG ^{a,1}, Stine HANGAARD ^a, and Ole HEJLESEN ^a

^a*Department of Health Science and Technology, Aalborg University, Denmark*

Abstract. Chronic obstructive pulmonary disease (COPD) is highly underdiagnosed. Early diagnosis is, therefore, essential to reduce costs and exacerbations and prevent disease progression. This calls for the development of a COPD case-finding tool. The present study aimed to develop a model for early identification of COPD with an eye to optimizing COPD case finding. The study was based on data from the US National Health and Nutrition Examination Survey (NHANES) dataset (2007-2012). For the analysis, 772 participants with spirometry defined COPD were included. Potential predictors for COPD (n=42) were extracted. The model was based on logistic regression, and the predictors were included using a stepwise forward selection. A five-fold cross-validation was used to train and validate the model. The predictors included age, gender, and pack-years of smoking. The model obtained an AUC of 0.71. In conclusion, such a model can be useful for identifying individuals that should perform post-bronchodilator spirometry to aid early identification of COPD.

Keywords. Airflow obstruction, COPD, prediction, diagnosis

1. Introduction

Spirometry remains the decisive marker in the diagnosis of chronic obstructive pulmonary disease (COPD). According to the Global Initiative for Chronic Lung Disease (GOLD), a post-bronchodilator forced expiratory volume (FEV1) / forced vital capacity (FVC) of <0.7 confirms the presence of COPD [1]. The diagnosis of COPD should also be based on an assessment of symptoms, comorbidities, and risk of exacerbation [1]. However, COPD remains highly underdiagnosed and misdiagnosed [2–4]. Globally, approximately 70% of subjects with COPD are undiagnosed [5].

Early identification of patients with COPD is warranted. Early diagnosis may reduce healthcare costs [6], reduce exacerbations, and prevent disease progression. [2,6,7]. However, no direct evidence supports that a broad spirometry-based screening of potential COPD patients will improve longitudinal health parameters [8]. Inline herewith, GOLD does not recommend spirometry screening, and advocates for active COPD case finding as an alternative [2].

Several studies have explored the characteristics of subjects with COPD who remain undiagnosed [9–13]. One may assume that a focused model based on potential characteristics of undiagnosed COPD have the potential to improve early identification of subjects with COPD. A previous study used 34 variables based on demographic, clinical, and health care resource utilization characteristics to identify subjects with

¹ Corresponding Author. E-mail: tk1@hst.aau.dk

undiagnosed COPD [14]. However, such a model is considered too complicated for implementation. Therefore, a simple and accurate case-finding tool is yet to be developed. Thus, the present study aimed to develop a simple model for early identification of COPD.

2. Methods

Data were extracted from the US National Health and Nutrition Examination Survey (NHANES) dataset (2007-2012). NHANES data are publicly available, and they include a variety of medical, nutritional, behavioural, and demographic data. The data is collected through examinations and interviews and are representative of the US population [15,16].

Data from a total of 23,433 participants were extracted from the NHANES database. A subsample of 1,564 participants whose FEV1/FVC ratio was below the lower limit of normal or below 0.70 were asked to repeat spirometry after inhaling a β 2-adrenergic bronchodilator medication to open up their airways. Participants below the age of 40 were also removed as COPD predominantly develops hereafter, and thus 772 participants remained for further analysis.

2.1. Stratification and feature extraction

The participants were stratified into participants with COPD ($n = 460$) and without COPD ($n = 312$) based on a post-bronchodilator FEV1/FVC < 0.7. A total of 42 potential predictors for COPD were extracted involving questionnaires related to age, anxiety, BMI, chest pain, colds recently, cough, days lost to wheezing, days of recreational activity, depression, difficulty walking up ten steps, difficulty walking a quarter-mile, dyspnea, education, emphysema, exposure to exhaust fumes, mineral dust, organic dust, and other fumes, fatigue, gender, general health condition, hay fever, height, hypertension, income, lung cancer, marital status, pack-years of smoking, phlegm, race, respiratory illnesses, shortness of breath, smoked 100 cigarettes, smoking currently, smoking in household, smoking amount in household, social difficulties, trouble sleeping, vigorous work, vigorous work amount, and waist circumference.

2.2. Model training and validation

The model was based on logistic regression, and the predictors were included using a stepwise forward selection. In each step, a predictor was included if it significantly improved the model compared to the previous step ($p < 0.001$).

A five-fold cross-validation was used to train and validate the model based on the included predictors. A concatenation of the model-estimated probability of COPD for each participant was then used to form a receiver operating characteristics (ROC) curve. A ROC curve displays the true positive rate against the false-positive rate at different thresholds. The area under this curve (AUC) is used for quantifying the general performance of the model, where the AUC ranges from 0 (no correctly classified targets) to 1 (perfect classification with no false positives).

3. Results

Table 1. Predictors included based on stepwise forward selection ($p < 0.001$)

Step	Predictor	p-value	AUC
1	Age	$1.21e^{-11}$	0.62
2	Gender	$5.39e^{-8}$	0.65
3	Pack years of smoking	$2.39e^{-4}$	0.71

The included predictors involved age, gender, and pack-years of smoking, which were then used to train and validate a logistic regression model. Figure 1 presents the ROC curve based on the model’s ability to differentiate between the 722 participants with and without COPD. The model obtained an AUC of 0.71.

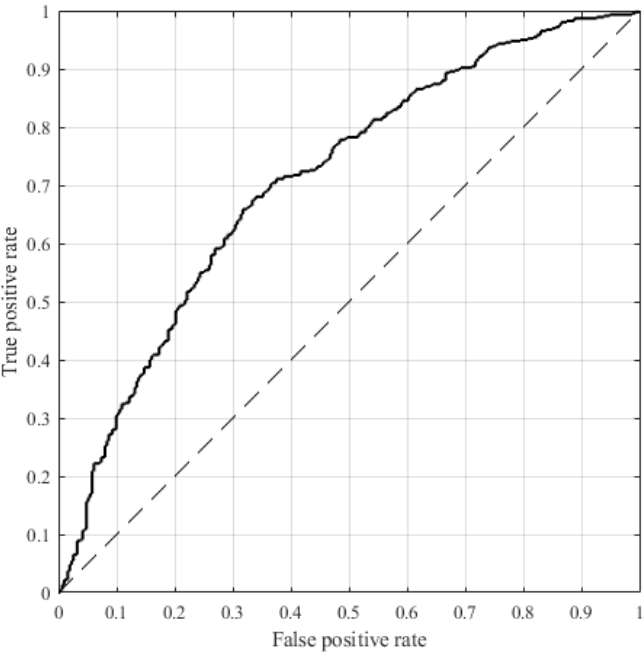


Figure 1. Receiver operating characteristics curve.

4. Discussion

The present study aimed to develop and validate a simple model to identify subjects with COPD. The model performance is considered acceptable as the model did not include any pulmonary function indicators. The model can, therefore, be used to identify subjects that should be considered for pulmonary function testing for early diagnosis of COPD.

The previous study by Moretz et al. [14] presented a stepwise logistic regression based on 34 variables, closely related to COPD, involving respiratory conditions, cardiovascular conditions, comorbidities, medications, hospitalizations, and patient characteristics. The model obtained an AUC of 0.75 with a sensitivity of 0.60 and specificity of 0.78. The present study presented a simple model based on age, gender, and pack-years, which is considered more likely to be successfully implemented. This model obtained an AUC of 0.71 and a sensitivity of 0.50 at a specificity of 0.78.

The present study was limited by basing COPD diagnosis solely on spirometry and not assessing symptoms, comorbidities, and risk of exacerbation.

5. Conclusion

A model approach, as described in the present study, can be used to identify individuals that should perform post-bronchodilator spirometry to aid early identification of COPD.

References

- [1] Global Initiative for Chronic Obstructive Lung Disease, Global Initiative for Chronic Obstructive, 2018. doi:10.1097/00008483-200207000-00004.
- [2] J.L. López-Campos, W. Tan, J.B. Soriano, Global burden of COPD, *Respirology* **21** (2016), 14–23. doi:10.1111/resp.12660.
- [3] J.B. Soriano, J. Zielinski, D. Price, Screening for and early detection of chronic obstructive pulmonary disease, *Lancet* **374** (2009), 721–732. doi:10.1016/S0140-6736(09)61290-3.
- [4] M.R. Miller, M.L. Levy, Chronic obstructive pulmonary disease: missed diagnosis versus misdiagnosis, *British Medical Journal* **351** (2015), h3021. doi:10.1136/bmj.h3021.
- [5] N. Diab, A.S. Gershon, D.D. Sin, W.C. Tan, J. Bourbeau, L.P. Boulet, et al., Underdiagnosis and Overdiagnosis of Chronic Obstructive Pulmonary Disease, *American Journal of Respiratory and Critical Care Medicine* **198** (2018), 1130–1139. doi: 10.1164/rccm.201804-0621CI.
- [6] D.W. Mapel, S.B. Robinson, H.B. Dastani, H. Shah, A.L. Phillips, E. Lydick, The direct medical costs of undiagnosed chronic obstructive pulmonary disease, *Value Health* **11** (2008), 628–636. doi: http://dx.doi.org/10.1111/j.1524-4733.2007.00305.x.
- [7] R. Rodriguez-Roisin, *Chronic Obstructive Pulmonary Disease Updated 2010 Global Initiative for Chronic Obstructive Lung Disease*, Global Initiative for Chronic Obstructive Lung Diseases Inc., 2016.
- [8] US Preventive Services Task Force, Screening for chronic obstructive pulmonary disease using spirometry: US Preventive Services Task Force recommendation statement, *Annals of Internal Medicine* **148** (2008), 529–534.
- [9] C.H. Martinez, D.M. Mannino, F.A. Jaimes, J.L. Curtis, M.L.K. Han, N.N. Hansel, et al., Undiagnosed obstructive lung disease in the United States associated factors and long-term mortality, *Annals of the American Thoracic Society* **12** (2015), 1788–1795. doi:10.1513/AnnalsATS.201506-388OC.
- [10] S. Hangaard, T. Kronborg, O. Hejlesen, Characteristics of Subjects With Undiagnosed COPD Based on Post-Bronchodilator Spirometry Data, *Respiratory Care* **64** (2019), 63–70.
- [11] S.C. Hvidsten, L. Storesund, T. Wentzel-Larsen, A. Gulsvik, S. Lehmann, Prevalence and predictors of undiagnosed chronic obstructive pulmonary disease in a Norwegian adult general population, *Clinical Respiratory Journal* **4** (2010), 13–21. doi:10.1111/j.1752-699X.2009.00137.x.
- [12] B. Lamprecht, J.B. Soriano, M. Studnicka, B. Kaiser, Determinants of Underdiagnosis of COPD in National and International Surveys, *Chest* **148** (2015), 971–985. doi:10.1378/chest.14-2535.
- [13] S.R. Santos, E.S. Lizzi, E.O. Vianna, Characteristics of undiagnosed COPD in a senior community center, *International Journal of Chronic Obstructive Pulmonary Disease* **9** (2014), 1155–1161. doi:10.2147/COPD.S49521.
- [14] C. Moretz, Y. Zhou, A.D. Dhamane, K. Burslem, K. Saverno, Jain G, et al., Development and Validation of a Predictive Model to Identify Individuals Likely to Have Undiagnosed Chronic Obstructive Pulmonary Disease Using an Administrative Claims Database, *Journal of Managed Care & Speciality Pharmacy* **21** (2015), 1149–1159. doi:10.18553/jmcp.2015.21.12.1149.

- [15] National Centers for Health Statistics, *National Health and Nutrition Examination Survey*, https://www.cdc.gov/nchs/nhanes/about_nhanes.htm.
- [16] G. Zipf, M. Chiappa, K.S. Porter, Y. Ostchega, B.G. Lewis, J. Dostal, National health and nutrition examination survey: plan and operations, 1999-2010, *Vital and Health Statistics* (2013), 1–37.

A Study of the Possibility of Detecting Pediatric Mild Developmental Delay Through a Serious Game: A Randomized Cluster Trial in Cambodia

Ho Heon KIM BSN^{a,1}, Hong Wook KWON BA^{a,1},
Sambath KAO MD, PhD^b, Hyuk-Sang SOHN PhD^c, Joohye KIM PhD^d,
Yejin CHOI BS^c, Doora Kim BS^c, Mirae Kang BS^c, Yu Rang Park PhD^{a,2}
^a *Department of Biomedical Systems Informatics Yonsei University College of Medicine, Korea*
^b *National Pediatric Hospital, Cambodia*
^c *Kyung Hee University, Korea*
^d *Busan Presbyterian University, Korea*
^e *Dobrain inc., Korea*

Abstract. Although early diagnosis of developmental delay is important, there are challenges in identifying cognitive status in developing countries because of limited human and financial resources to perform diagnostic tests. Moreover, diagnosis stability of developmental delay in children using neuropsychological tests (NPTs) can remain unsettled. The aim of this study is (1) to verify the effectiveness of a serious game (DoBrain), (2) to identify existing inconsistencies between NPTs, and (3) to explore the potential of the serious game as a complement to diagnostic tools. Eligible children who had completed results of NPTs were selected (n=119/235; 116/235; case, control). With these children's scores, we performed the Mann-Whitney U test to investigate the effectiveness of the serious game by comparing the improvement of scores in both groups. Among the participants, we additionally selected a case group to identify the potential of the serious game for detecting mild developmental delay. Using the results of the CGI-S as a baseline, we defined the participants whose scores indicated more than mild illness (≥ 2 points) in at least one area as the suspected group. The score improvement related to memory in case group was greater than that of the control group ($p < 0.05$). Furthermore, four of the NPTs were not inconsistent, and the sensitivity/specificity of DDST-II was the highest score considering CGI-S results as the ground truth (0.43; 0.96). Additionally, games measuring discrimination, velocity, memory, and spatial perception showed statistical significance ($p < 0.05$). This study verifies that the serious game can help specific cognitive areas and suggests that the serious game could be used as a low-cost and unconstrained spatiotemporal alternative to NPTs.

Keywords. Digital therapeutics, digital phenotyping, serious games, inter-observer variability

1. Introduction

Developmental delay in early childhood tends to deteriorate or spread into other psychological disorders gradually. For this reason, early detection of developmental delay at an early age is important [1]. However, conventional neuropsychological tests (NPTs) for developmental delay have two issues: the inter-observer and intra-observer variability [2] and lack of accessibility to the diagnosis of developmental delay due to cost or time demands [3]. This problem is more apparent in developing countries where access to health service is relatively difficult [4]. The purpose of this study is (1) to verify the effectiveness of a serious game, (2) to investigate the discordance of NPTs and (3) to identify the possibility of mild developmental delay in Cambodia by using a serious game.

2. Method

2.1. DoBrain and Population recruitment

This study was based on a multi-centre, two-arm, cluster-randomized controlled trial with the first endpoint to identify the effectiveness of serious mobile game (DoBrain) for cognitive development of children in Cambodia. DoBrain is a serious game based on an animated cartoon and is a smart device-based application certified for child suitability. We chose this game as an intervention tool because this app showed a high retention rate (42.5%) for 30 days compared to an 8% retention rate of general health apps [5].

In a clinical trial, we recruited a total of 403 children from Xavier School (Sisophon) and Mirero School (Phnom Penh). The participants were males and females between three and nine years old. By choosing each four classes randomly to assign into case and control group, we conducted a case-control study with a case group of 204 participants and a control group of 199 participants, which lasted from March 11 to July 31 in 2019. We assessed each group with pre-NPTs to appraise the children's overall cognitive function, executive function, memory, and attention. These included the following: (1) the clinical global impression – severity scale (CGI-S), (2) the Denver developmental screening test, second edition (DDST II), (3) the Cambridge neuropsychological test automated battery (CANTAB), and (4) the behaviour rating inventory of executive function, second edition (BRIEF-II). We guided study participants through the process of downloading, starting, and using DoBrain. Participants used provided software 2 days a week for 20 weeks. Finally, we again assessed each group using the same NPTs to measure the participants' overall cognitive function. Ethical clearance was obtained from the Ministry of Health National Ethics Committee for Health Research in Cambodia (Number: 212) before participants were recruited.

2.2. Selection of eligible users and analytic methods

Of the total 403 participants, we included those who completely underwent the four NPTs as eligible users ($n=236$). Further, we excluded the participants who did not have two pre- or post-tests (BRIEF-II, CANTAB) to evaluate the effectiveness of the serious game ($n=1$). With these participants ($n=235$), we figured out the effectiveness of this

app (Table 1) and identified the inconsistencies between NPTs (Figure 1). Furthermore, only the case group were selected to identify the potential of the serious game for detecting mild developmental delay because they had experience with DoBrain (n=119) (Table 2).

Because the line between normal and abnormal in NPTs is ambiguous, we defined suspect results of tests for developmental delay, as follows, in order to identify inconsistency between NPTs for developmental delay: (1) CGI-S: cases where a clinician with a PhD in pediatric psychology gave severity scores ranging from mild to severe (two or more points) in at least one cognitive subfield; (2) DDST-II: cases that received a caution or delay score in at least one area in each cognitive subfield; (3) CATNAB: cases with scores less than 3% on at least one sub-test; and (4) BRIEF-II: cases with at least one area with a score more than 65 converted to a t-score considered clinically significant. At this cutoff, we established a percentage to classify children as low growth, that is, less than 3% of the growth curve of normal child development. In addition, we considered the CGI-S results as a foundation for comparing the consistency of other NPTs because CGI-S is a clinician-determined measure that takes into account all available information about patients.

Finally, we conducted the Mann-Whitney U test to investigate the effectiveness of the serious game by comparing the improvement of scores in both groups (Table 1). For each test, the suspect results were visualized by binary coding: 1 for suspect and 0 for normal (Figure 1). Further, to identify the potential to classify the CGI-S results by game scores, we examined the difference of game scores in each group by performing a t-test from 1 chapter to 63 chapters (Table 2).

3. Results

Case and control groups did not show a significant difference in age and proportion of sex (their mean age is 7.0, 6.9 years, respectively; $p=0.620$; 0.613, ratio of male is 69/116 (59.5%), 71/119(59.6%)). After intervention, PRMPCI and SWMBE486 scores related to memory area were significantly improved in both groups ($p<0.05$). Furthermore, the score improvement in the case group was greater than that of the control group ($p<0.05$) (Table 1). MOTML related to attention, and psychomotor speed showed no significant difference between before and after intervention in case group ($p=0.556$).

Table 1. Comparison with pre and post-test of Case and Control group in BIREF-II test and CANTAB test

Outcomes	Case (n=119)		Control (n=116)		Median difference between groups†	p-value
	Pre test	Post test	Pre test	Post test		
BRIEF (Median, IQR)						
Emotional control	56.0 (27.0)	52.0 (13.5)	66.0 (20.5)	59.0 (21.3)	-6.0	0.025
Inhibit	55.0 (19.0)	52.0 (13.0)	57.0 (16.3)	53.0 (13.0)	-4.0	0.002
Initiate	53.0 (18.0)	53.0 (9.0)	56.0 (9.0)	53.0 (11.0)	-7.0	<.001
Organization of material	49.0 (16.0)	52.0 (13.5)	52.0 (16.8)	49.0 (16.0)	0.0	0.120
Plan/Organization	51.0 (18.0)	49.0 (12.0)	56.5 (15.0)	49.0 (13.0)	-8.0	<.001

Task monitor	51.0 (19.0)	50.0 (15.0)	54.0 (17.0)	49.0 (10.0)	-3.5	<.001
Working memory	53.0 (24.5)	52.0 (14.5)	57.0 (14.0)	52.0 (14.0)	-3.0	0.016

CANTAB (Median, (IQR))

MOTML (latency time)	777.6(231.8)	799.1(235.2)	867.8(340.4)	791.0(273.5)	-69.2	0.022
PRMPCI (correct patterns)	66.7(33.3)	91.7(33.3)	75.0(33.3)	83.3(27.1)	-21.1	0.002
PRIFMDR (reaction time)	515.0(135.0)	530.0(111.5)	514.0(104.4)	510.8(88.1)	-13.7	0.122
PRFMDMT (movement time)	270.0(102.5)	267.5(72.6)	303.5(95.8)	305.0(70.6)	-3.2	0.379
SSPFSL (correct sequences)	4.0(2.0)	4.0(2.0)	3.0(1.0)	4.0(2.0)	0.4	0.200
SWMBE468 (error times)	25.0(8.5)	22.0(8.0)	24.0(6.3)	24.0(6.0)	2.5	0.010

†Estimated between-group difference (intervention –control) after the intervention (post-test), for pre-test values

The CGI-S result showed that 200 participants were classified into the normal group, and 35 participants were classified into the suspect group (85.1% and 14.9%, respectively). With these participants, inconsistencies were seen between the results of the four NPTs (Figure 1). In addition, using the results of CGI-S as a baseline, there were variabilities in the sensitivity and specificity of DDST-II, CANTAB, and BRIEF-II (sensitivity, specificity: 0.43, 0.96; 0.34, 0.67; 0.89, 0.49; respectively).



Figure 1. Inconsistency between four neuropsychological tests ([A] heatmap and [B] Venn diagram)

In the case group, chapters of DoBrain 4, 5, 22, 23 and 24, showed statistically significant differences in average game scores of the normal and the developmental delay group. In particular, the average of game scores for measuring discrimination, spatial perception and inference, organizing, memory and velocity perception areas in chapters 4, 5, 22, 23, and 24 were significantly higher than suspected group ($p<0.05$).

Table 2. Chapters in DoBrain showed statistically significant differences based on CGI-S results

	Normal Group (n=105)	Suspected group of developmental delay (n=14)	p- value
Chapter 4: Discrimination, (median (IQR))	10.5 (0.6)	9.2 (1.1)	0.048
Chapter 5: Spatial perception, (median (IQR))	7.9 (3.3)	6.9 (2.2)	0.011
Chapter 22: Inference (median,(IQR))	8.9 (2.2)	8.1 (1.3)	0.037

Chapter 23: Organizing, (median (IQR))	8.2 (1.4)	7.3 (1.3)	0.008
Chapter 24: Memory, (media, (IQR))	10.2 (1.6)	9.6 (1.4)	0.017
Chapter 24: Velocity perception, (media, (IQR))	8.2 (1.4)	7.1 (1.1)	0.006

* The number of children who played each game is slightly different for each game chapter

4. Discussion

The first major finding of this study is that this serious game can improve the memory area of children in Cambodia. To identify the effectiveness of this intervention, we adjusted the effect of the natural development process by comparing differences in scores between the case and the control group. The scores of BRIEF-2 did not change; however, BRIEF-2 is based on the scores reported by teachers, so we considered it to be the result of observer-variability in the BRIEF-II test.

Second, the conventional tests were not consistent and varied in criteria for suspected mild developmental delay, even in the same child. Although the purposes of these tests are to screen for developmental delay in children or to measure cognitive level, this inconsistency may result from measurements of heterogeneous areas. For example, BRIEF-II can measure eight cognitive areas in detail; however, DDST-II can measure only four areas.

Lastly, we found games where the average of game scores differed between the two groups related to CGI-S results. In particular, the discrimination, inference, velocity, memory and spatial perception categories within the game showed high relevance.

This study has a limitation that cognitive levels of children were over or underestimated because some children were suspected of having mild developmental delay by comparing their score to that of other children (under 3%). To make our findings more robust, further research with clear diagnosis data is needed.

In conclusion, this study verified that a serious game could help in the improvement of specific cognitive areas and showed that inconsistencies exist among four conventional NPTs. It also identified the potential for serious games to be used as a low-cost, unconstrained spatiotemporal NPT in low- and middle-income countries.

References

[1] D.L. Robins, D. Fein, M.L. Barton, J.A. Green, The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders, *Journal of autism and developmental disorders* **31**(2) (2001), 131-44.

[2] S.M. Gore, Assessing clinical trials--between-observer variation, *British medical journal (Clinical research ed.)* **283**(6283) (2001), 40.

[3] D.L. Morelli, et al., Challenges to implementation of developmental screening in urban primary care: a mixed methods study, *BMC pediatrics* **14**(1) (2014), 16.

[4] S. Saxena, et al., Resources for mental health: scarcity, inequity, and inefficiency, *The Lancet* **370**(9590) (2007), 878-89.

[5] G.H. Tison, et al., Achieving High Retention in Mobile Health Research Using Design Principles Adopted from Widely Popular Consumer Mobile Apps, *Circulation* **136**(suppl_1) (2017), A21029-A.

Current Use of Sensor-Based Measurements for Paraplegics: A Literature Review

Nils-Hendrik BENNING^{a,1}, Niclas HAGEN^a and Petra KNAUP^a

^a *Institute of Medical Biometry and Informatics, Heidelberg University, Germany*

Abstract. Sensors are used in many fields to measure physical phenomena, often on mobile persons. Paraplegia is a disease with a massive impact on the ability to move, so patients show changes in walking behaviour or are even wheelchair users. It is unclear how sensors can be used with paraplegics to generate valid data for research. In the ParaReg project, such data shall be integrated with a medical register. In this paper, we elaborate current approaches for sensor measurements that might serve as an additional data source for ParaReg with the help of a literature review. We queried IEEE Xplore and MEDLINE to find publications that describe the current use of sensors for patients with paraplegia. The retrieved publications were screened for eligibility by two reviewers independently. To ensure reproducibility, we conducted an initial alignment. All publications were assessed based on their abstract. Inclusions were analysed for their major topic, and we proposed categories, which were statistically described afterwards. We retrieved 685 publications in total and included 287 publications in our analysis. The categories we found are: “diagnostic sensor tools for clinical environments”, “seating position analysis”, “functional electrical stimulation (FES) / neuroprosthesis control”, “seating position analysis”, “control systems for devices” and “assessment of physical activity”. “FES / neuroprosthesis control” shows most publications and the highest publication rate in history, followed by “assessment of physical activity”.

Keywords. Sensor, paraplegia, spinal cord injury, literature review.

1. Introduction

The use of sensors to measure physical functions is an evolving field. Such measurements can be used to observe various physical phenomena or even to control technical systems [1]. One example of an observational measurement is the assessment of patient postures. Typical methods and solutions in this field are suited for healthy or mobile persons. But paraplegia is a disease which limits the physical functions of patients significantly [2]. So, these methods might not be appropriate for paraplegics because they show unwanted different walking patterns [3] or even use a wheelchair.

In the ParaReg project, we want to find out how sensor data can contribute to the database of medical registers, to enrich the medical information value. With this

¹ Corresponding Author, Nils-Hendrik Benning, Institute of Medical Biometry and Informatics, Heidelberg University, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany; E-mail: nils.benning@med.uni-heidelberg.de.

literature review, we want to elaborate on current approaches for sensor measurements in care and research that might be used in ParaReg.

2. Methods

We systematically reviewed original works in IEEE Xplore and MEDLINE databases. Our research question for this review is: for which purpose were sensors used for patients with paraplegia. We accepted reports on experiments or other kinds of presentations, like developed prototypes or concepts. For our study, we considered articles from all years in English and German language. The search was performed on the 1st June 2019 and updated on 1st July 2019 using the following search strategy, exemplarily presented based on the PubMed query.

```

("sensor"[Title/Abstract]
 OR "sensors"[Title/Abstract]
 OR "gyroscope"[Title/Abstract] OR "gyroscopes"[Title/Abstract]
 OR "accelerometer"[Title/Abstract] OR
"accelerometers"[Title/Abstract] OR "IMU"[Title/Abstract]
 OR "inertial measurement unit"[Title/Abstract]
 OR "inertial measurement units"[Title/Abstract]
 OR "inertial sensor"[Title/Abstract] OR "inertial
sensors"[Title/Abstract] OR "GPS"[Title/Abstract] OR "Global Positioning
System"[Title/Abstract] OR "Fitness Trackers"[Mesh]
 OR "Fitness Tracker"[Title/Abstract] OR "Fitness
Trackers"[Title/Abstract] OR "Activity Tracker"[Title/Abstract] OR "Activity
Trackers"[Title/Abstract] OR "Activity Monitoring"[Title/Abstract])
AND
("Spinal Cord Injuries"[MeSH Terms] OR "SCI"[Title/Abstract]
 OR "Spinal Cord Injury"[Title/Abstract]
 OR "Spinal Cord Injuries"[Title/Abstract]
 OR "paraplegia"[MeSH Terms] OR "Paraplegia"[Title/Abstract])

```

The query consisted of two main parts, linked by an AND condition. The first part queried for sensors. The keyword “sensor” was extended by common sensor forms, which might be used instead of the general term “sensor”. They were added with OR conditions, so the query became broader and not narrower. The second part of the query was specific for the disease. Because of the ambiguous use of language, we included both paraplegia and spinal cord injuries.

The search strategy was transferred to the IEEE syntax in order to query the second database in the same manner. Besides these queries, we did not use any filters or limits.

The retrieved articles were merged into a single reference list, and then duplicates were eliminated. The screening started based on abstracts, which was done by two reviewers independently. Eligibility of an abstract was assessed on the fact if it was assignable to a type of sensor-based measurement for paraplegics. Furthermore, we used the following exclusion criteria:

- E1 – Studies not published in English or German
- E2 – Publications not related to patients with paraplegia or spinal cord injury
- E3 – Reported exclusively on animals
- E4 – Not presenting original work in the field
- E5 – Presentation of work in the field of Biochemistry

We conducted an abstract-based initial alignment with both reviewers. We measured the agreement of the decision for either “include, relevant publication” or one of the exclusion reasons (E1-E5). Afterwards, the agreement of the reviewers was assessed with Cohen Kappa; values above 0.75 were accepted [4]. Disagreement during the initial assessment was solved by an intensive discussion of the article’s abstract and on content. After the initial assessment, all remaining publications were split and each of them assigned to one reviewer. For the included publications, the reviewers began to develop categories, describing the kind of sensor-measurement in the use case. This development was done cooperatively and inductively. Each inclusion was assigned to the best fitting category. In cases where there was no appropriate category defined, the reviewers discussed different approaches before introducing a new one.

The included and categorized articles were exported to a separate table and analysed statistically.

3. Results

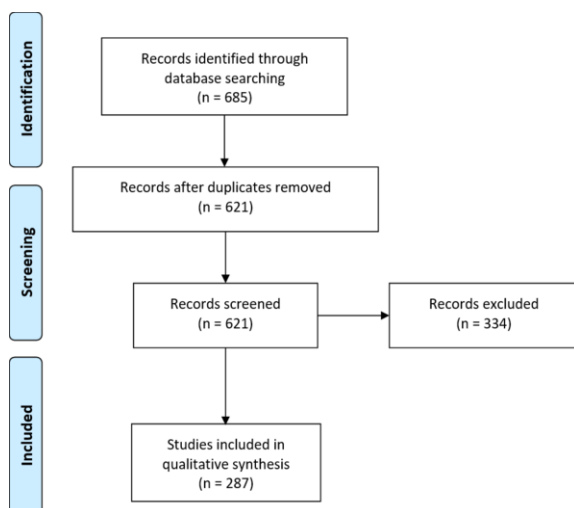


Figure 1 - Flow Chart of the screening process (adopted from [5])

In total, we retrieved 685 publications. There were 621 distinct ones without duplicates. During the initial assessment, we achieved a Cohen Kappa of ~ 0.85 , which was accepted due to the given threshold of 0.75. We excluded 334 publications during the screening. So, in total, we included and thus classified 287 publications, which can be found in the published dataset [6].

The analysis of the abstracts resulted in five categories. These categories have been assigned to the included publications. The categories and the share of publications in them are:

C1 (9,1 %) – Diagnostic sensor-based tools, which are suited only to clinical environments or study centres (non-mobile)

C2 (9,1 %) – Seating position or pressure distribution analysing in wheelchairs

C3 (50,9 %) – Control systems for FES devices (functional electrical stimulation) or other neuroprostheses

C4 (11,5 %) – Human control interfaces for devices (e.g. personal computers)

C5 (19,5 %) – Focus on the assessment of physical activity

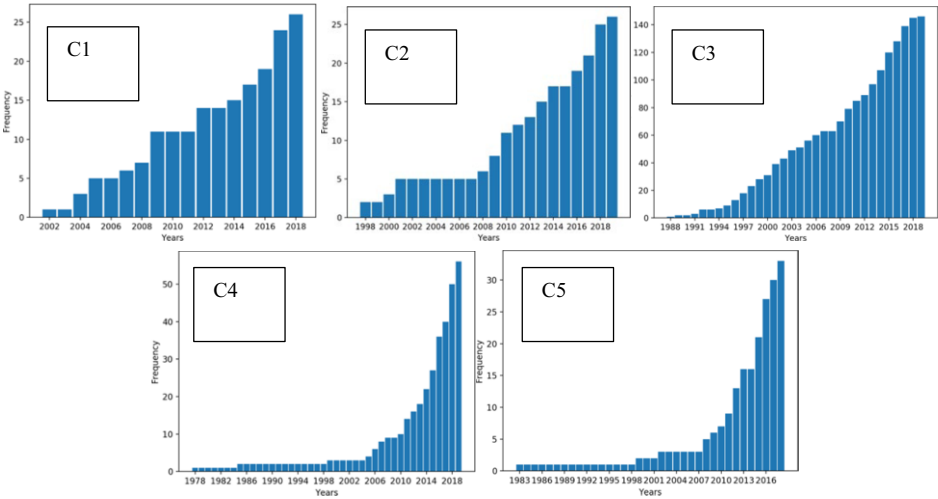


Figure 2 illustrates grafically the cumulative publication activity for the five categories.

To analyze if the found publications are also an evolving topic, with recent publications and ongoing activity, we analyzed the publication history. Figure 2 shows the cumulative publication activities. C3: “FES / neuroprosthesis control” (Figure 2) only has very few periods where no new publications appeared, and it increased rather steadily. In contrast, “assessment of physical activity” (C5) did not have new publications for 14 years in a row (see Figure 2). “Diagnostic sensors for clinical environments” (C1), “seating position/pressure analysis” (C2), “control systems for devices” (C4) and especially “assessment of physical activity” (C5) have a stronger increase in 2018. “FES / Neuroprosthesis control” (C3) shows a comparable increase in 2016/2017.

Table 1 - Total number of publications in the given topics

Measure	C1	C2	C3	C4	C5
Sum	26	26	146	33	56
Share	9.1 %	9.1 %	50.9 %	11.5 %	19.5 %

4. Discussion

All included publications could be assigned to one of the five categories. “FES / neuroprosthesis control” (C3) seems to be the field of research with the longest history and richest experience. Younger categories are “control systems for devices (such as personal computers)” (C4), “diagnostic sensors for clinical environments” (C1) and “seating position/pressure analysis” (C2).

First publications appeared remarkably close to each other in the categories “diagnostic sensors for clinical environments” (C1), “seating position/pressure analysis” (C2) and “control systems for devices” (C4) (years 1998-2002). Improved availability for sensor technology could explain this. The first publications in “FES /

neuroprosthesis control” (C3) appeared a decade earlier and continuously increased, which suggests a higher maturity of C3 compared to the other fields. First publications on “assessment of physical activity” (C5) appeared in 1978; however, there was a long stagnation of the publication counts until 2005.

The insights of this systematic review have to be interpreted with care. Even though the search strategy was very generic, the used keywords might always lead to selection bias, and thus the results cannot claim to be exhaustive. However, by using one medical database and one technically focussed database, combined with a generic search strategy, we could identify a reasonable number of relevant articles and corresponding categories.

5. Conclusions

The use of sensors for paraplegics is not new (first publications from 1978), but it is still a highly relevant topic, as publications rates are even increasing within five categories during the last years. This encourages sensor-based data generation in one of the categories found for integration into the ParaReg register, to support medical research with a more extensive database.

6. References

- [1] D. Dias and J. Paulo Silva Cunha, Wearable Health Devices-Vital Sign Monitoring, Systems and Technologies, *Sensors (Basel)* **18** (2018).
- [2] S. Franz, S.C. Kirshblum, N. Weidner, R. Rupp, and C. Schuld, Motor levels in high cervical spinal cord injuries: Implications for the International Standards for Neurological Classification of Spinal Cord Injury, *J Spinal Cord Med* **39** (2016), 513–517.
- [3] D. Schließmann, M. Nisser, C. Schuld, T. Gladow, S. Derlien, L. Heutehaus, N. Weidner, U. Smolenski, and R. Rupp, Trainer in a pocket - proof-of-concept of mobile, real-time, foot kinematics feedback for gait pattern normalization in individuals after stroke, incomplete spinal cord injury and elderly patients, *J Neuroeng Rehabil* **15** (2018), 44.
- [4] W. Greve, D. Wentura, H. Gräser, and U. Schmitz, *Wissenschaftliche Beobachtung: Eine Einführung*, 2nd Edition, Beltz, Weinheim, 1997.
- [5] D. Moher, A. Liberati, J. Tetzlaff, and D.G. Altman, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *PLoS Med* **6** (2009), e1000097.
- [6] N.H. Benning, N. Hagen, and P. Knaup, Sensor-Based Measurements in Paraplegia: Classified References from a Systematic Review, <https://doi.org/10.11588/data/JRVJGN>, *heiDATA*, **V1** (2020).

Detection of Psychomotor Agitation Pattern from Motion Sensor Data in a Living Environment of a Patient with Dementia

Corinna MIELKE^{a,1}, Rasmus ANTONS^a and Reinhold HAUX^a

^a*Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Braunschweig, GERMANY*

Abstract. Agitation is a symptom of many mental illnesses such as dementia. The aim of this work is to detect psychomotor agitation pattern in the living environment of a patient with dementia. To reach this goal, we built a smart home environment to collect the motion data of our patient. These data can be used to show the single movements of a person between different rooms in this apartment. The interpretation of these movements allows us to conclude if movements are normal or look like psychomotor agitation, in particular wandering around. This information might be useful for further analysis and interpretation. Additionally, it might be possible that the movement can give an indication of possible mood and behavioural changes in the patient. So it should be possible to create a long-term living environment that enables residents to live a long and independent life in their own home.

Keywords. Mental illnesses, dementia, behaviour monitoring and interpretation, ambient assisted living, health-enabling and ambient assisted technologies, home automation technology, smart home

1. Introduction

Agitation describes a neuropsychiatric symptom which is characterized by pathologically increased, unproductive motor activities and the feeling of being driven and agitated. [1] According to this definition, agitation manifests itself through excessive motor activities such as nesting, reaching around, constant undressing, trembling or pointless wandering. The basic characteristic of agitation is that the hyperactive movements are carried out hastily, erratically and aimlessly. In particular, this symptom can be found in mental illnesses such as dementia or depression. [1]

The review by Khan et al. [2] gives a brief overview of the existing studies that use sensors to detect agitation in patients with dementia. They conclude, “*that actigraphy shows some evidence of correlation with of agitation and [...] multi-modal sensing has not been fully evaluated for this purpose*” [2]. Therefore, the research question of this paper is, if it is possible to detect *incidences* behaviours such as aimless wandering in the home with home automation technologies and health-enabling and ambient assistive technologies (HEAAT). This question is based on the work of Haux et al. and

¹ Corresponding Author: Corinna Mielke, Peter L. Reichertz Institute for Medical Informatics (PLRI), University of Braunschweig – Institute of Technology and Hannover Medical School, Mühlenpfordtstr. 23, D-38106 Braunschweig; E-mail: corinna.mielke@plri.de.

their conclusion that “*HEAAT remain an important field for future health care and for interdisciplinary research*” [3]. In addition, Mielke et al. describe in [4] the first example of an apartment as a diagnostic and therapeutic area. Following this example, it is possible to build up several modern apartments equipped with information and communication technology for medical care. They used home automation technologies and HEAAT to build this apartment, and as a result, they presented data from various types of sensors. Therefore, our goal is to build a Smart Home for medical care with the function to detect the psychomotor agitation pattern from motion sensor data in a living environment of a patient with dementia.

2. Method

In cooperation with Baugenossenschaft Wiederaufbau eG we equipped a housing unit in a building complex of assisted living with a home automation system, which can record and analyze selected data of the resident. Our first step to build this system was to develop a technical concept and a design on how to equip the apartment with sensors and smart home technologies. We decided to use the home automation system “provedo” with flat cables to minimize expenditure and cost for retrofitting. This home automation system was built by former company PROVEDO GmbH, and retrofitting was done by Baugenossenschaft Wiederaufbau eG. We equipped the apartment with a mini-computer and the necessary software to collect and analyze home automation data. In the following, we describe some details of this home automation system, the algorithm and the graphical user interface we developed for the detection of movement agitation patterns from motion sensor data in a living environment.

2.1. Details of the home automation system – sensors, data, resident

The apartment was equipped with different sensors. The installation includes motion sensors for activity detection, power sensors to collect the energy consumption, temperature and humidity sensors for automatic control of the heating system as well as contact sensors to detect the opening and closing of windows and doors. Therefore, the resident can experience several comfort functions like an automatic heating system, a chip-controlled door lock for keyless opening, a floor lighting as a night light, a stove and window control display and a stove monitoring. Additional to these comfort functions, the data collection runs autonomous and without any intervention of the resident. The sensors we used in the system are not intrusive and inconspicuous. PROVEDO offered a house bus system (see figure 1), which could easily be installed in existing apartments. Flat ribbon cables were used which could be placed under the wallpaper. The sensors and actuators are small devices that were mounted on the wall or ceiling like a light switch or smoke detector and thus could be integrated inconspicuously into the entire electrical installation of the home. Figure 1 shows the positions of the sensors in the apartment. The data recorded by the sensors are movement data within the apartment, the power consumption of the apartment, power consumption of the stove, room temperature and humidity, window status (open/closed) and access control. For the detection of psychomotor agitation, we used data recorded in the retrofitted apartment by a male subject aged 85+ and diagnosed with dementia between October 2016 and October 2018. The subject’s guardian was

informed about the project details and a signed consent form for the collection and use of data is available.

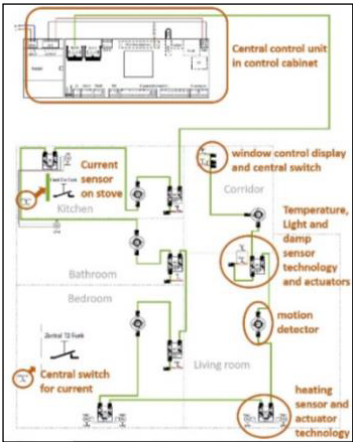


Figure 1. Schematic for the installation of sensors in the apartment with designation of individual components.

2.2. Algorithms for detection of movement agitation pattern

A relevant evaluation related to the behaviour of a person is the detection of inner restlessness and agitation. This evaluation is possible with the help of the analysis of all movement events over a certain period. In doing so, coherent movement sequences are identified, documented and evaluated. Connected motion sequences result from observation of individual motion events between rooms and their temporal distances. These movement sequences can be observed, recorded and evaluated by monitoring movements through individual rooms of the subject's home. In particular, the duration of an agitation phase and the frequency of room changes are of interest.

Motion sensors on the ceiling of each room are the basis for movement detection. They generate an event every time they detect motion, which is stored as the number of events per second. We only attempt to detect movements when one person is present in the apartment. The first step is to create a list of the resident's location for every time of day. Because the detection radius of sensors in neighbouring rooms overlap, it is possible to detect one person as being present in two rooms simultaneously. When no sensor detects motion, the resident has either left the residence or is rather motionless in the current room. We assume the resident has left the apartment if no motion is detected for 30 minutes, and the last room was the corridor. The following list with the current timestamp and the corresponding room in the residence can be derived from the data.

```
[ [1498953600000, ["bathroom"]],
  [1498953955000, ["bathroom", "corridor"]],
  [1498954003000, ["corridor"]],
  [1498954037000, ["living room"]],
  [1498954044000, ["living room", "bedroom"]],
  [1498954091000, ["bedroom"]],
  ...]
```

The list of all movement sequences is composed of individual movements, which are aggregated to a complete sequence if on the last individual movement the occupant stays at the same place for at least three minutes. A sequence of movements is marked

as critical if it lasts more than three seconds and contains the same room more than three times none consecutively. It is also marked critical if it contains impossible room changes, i.e. a direct change between two non-neighbouring rooms because this could indicate that more than one person is present.

3. Results

The data analysis results in a list of movement sequences. Due to its type and quantity of information contained, this list is now the basis for assessment and evaluation of individual movement sequences with regard to inner restlessness and agitation. For the research apartment, figure 2 shows the general overview of the evaluation for the movement data. The schematic floor plan with the possible movement points for orientation and better visual comprehensibility of movement sequences in the apartment is shown. The selected period is also displayed as well as the general summary for all found motion sequences and the motion sequences that were marked as critical through the evaluation for inner restlessness. In addition, this illustration also shows a normal motion sequence from September 26, 2017, with a defined start and end of the motion and the list with single movement sequences detected in the selected period. The motion sequences of this list that are marked as critical in the sense of inner restlessness are shown in red. Figure 2 shows a passage to the bathroom with a duration of 1 minute and 16 seconds.

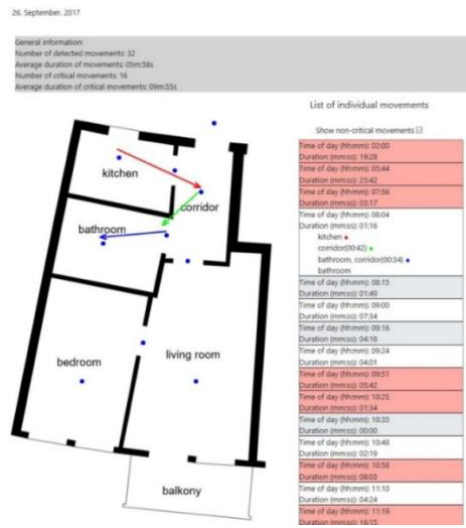


Figure 2. Controlled and targeted movement - a normal motion sequence with a defined start and end (Example from September 26, 2017, blue dots - motion points for single and neighbouring rooms, a motion is marked with an arrow in the floor plan and the same colour dot in the list of motions).

In contrast, figure 3 shows an accumulation of movement on the same day that has been evaluated as critical or conspicuous. This motion sequence is characterized as critical because each stay per room has lasted less than three minutes and contains the same room more than three times none consecutively.

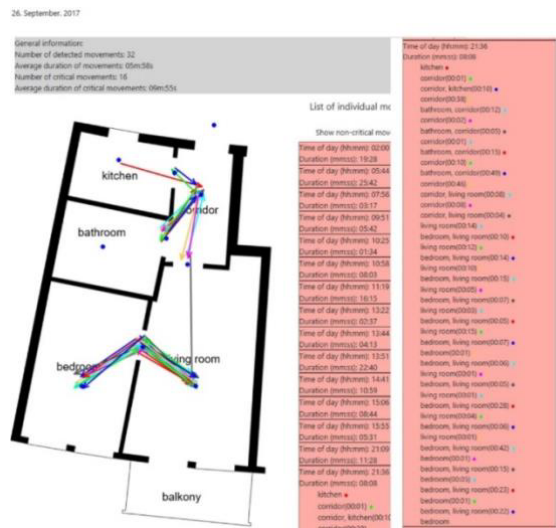


Figure 3. Agitation movement – movement sequence with many room changes about a duration of 8 minutes (Example from September 26, 2017, blue dots - motion points for single and neighbouring rooms, a motion is marked with an arrow in the floor plan and the same colour dot in the list of motions).

4. Discussion

The movement sequences marked as conspicuous or critical can give an indication of possible changes. The accumulation or temporal shift of such movement sequences could be indicators of a changing state of health in patients with dementia.

With the knowledge gained from this, we can develop systems that detect changes in lifestyle at an early stage. This enables relatives, caregivers or -providers, and doctors to intervene in a supportive manner in time. Thus, it should be possible to create a long-term living environment that enables residents to live a long and independent life in their own homes. In future work, we want to refine the algorithm, so that it is possible to improve the statements regarding agitations. Furthermore, we consider how a suitable study plan could look like in order to investigate to what extent agitation episodes can possibly be reduced with the help of the developed algorithm.

References

[1] M. Garriga, I. Pacchiarotti, S. Kasper, et al., assessment and management of agitation in psychiatry: Expert consensus, *World J Biol Psychiatry* **17**(2) (2016), 86-128.
[2] S.S. Khan, B. Ye, B. Taati, A. Mihailidis, Detecting agitation and aggression in people with dementia using sensors-A systematic review, *Alzheimers Dement.* **14**(6) 2018, 824-832.
[3] R. Haux, S. Koch, N.H. Lovell, et al., Health-Enabling and Ambient Assistive Technologies: Past, Present, Future, *Yearb Med Inform* **25** (2016), S76-S91.
[4] C. Mielke, T. Voss, R. Haux, Residence as a Diagnostic and Therapeutic Area-A Smart Home Approach, *Studies in health technology and informatics* **238** (2017), 92-95.

ECG-Quality Assessment of Dry-Electrode Cooperative Sensors

Michael RAPIN^a, Josias WACKER^a, Damien FERRARIO^a, Etienne HAENNI^a,
Stephan DASEN^a, and Olivier CHÉHELAT^{a,1}

^a CSEM SA, Neuchâtel, Switzerland

Abstract. Classical approaches for measuring high-quality ECG require the use of gel electrodes and individually shielded cables, which limit patient comfort, especially in long-term use. We recently introduced a novel sensing architecture—so-called cooperative sensors—that allow the use of active dry electrodes connected by two unshielded wires. The aim of this work is to qualitatively evaluate an ECG recorded with a dry-electrode cooperative-sensor system. To that end, preliminary observations were made on three healthy subjects. The ECGs were concurrently recorded with cooperative sensors and a gold-standard 12-lead ECG device during a stress test on a stationary bicycle. First experimental measurements demonstrated the reliability of the approach for a wearable 12-lead ECG monitoring system tested in real settings.

Keywords. Wearable sensors, dry electrodes, electrocardiography (ECG), multi-lead, active electrodes, cooperative sensors, long term monitoring.

1. Introduction

For many years the electrical activity of the heart has been measured with adhesive gel electrodes connected with shielded cables to a centralized electronic device used to record the electrocardiogram (ECG) signals [1]. One of the limitations that blocked the market entry of alternative wearable solutions—such as textile electrodes—was their signal quality that was not demonstrated as sufficient for medical use [2]. In order to solve this issue, we developed and patented active dry electrodes called “cooperative sensors” [3]. Cooperative sensors rely on voltage-sensing analog frontends which are directly placed onto the skin, allowing to maximize the input impedance and to reduce the sensitivity of the system to electromagnetic disturbances [4]. In addition to replacing glued gel electrodes with more comfortable, glue-free dry electrodes, cooperative sensors do not need shielded cables but work with a non-shielded two-wire bus which makes their integration in a garment simpler and further increases the wearing comfort of the system.

Since cooperative sensors use a different transducer (dry instead of gel electrodes) and a different way of application (pressure from a tight-fitting vest instead of an adhesive), tests cannot be limited to compliance to medical standards (e.g., IEC60601-2-47 or IEC60601-2-25), but require a qualitative assessment in a real setting. This work presents such an assessment.

¹ Olivier Chételat, Corresponding author, CSEM, Rue Jaquet-Droz 1, 2000 Neuchâtel, Switzerland; E-mail: olivier.chetelat@csem.ch.

2. Methods

The prototype depicted in Fig. 1 uses ten cooperative sensors embedded in a vest to measure 12 ECG leads [5]. The cabling, connecting the sensors together, consists of a bus of two non-shielded wires integrated into the textile of the vest. Sensors are located according to the Mason-Likar placement [6]. They are attached to the vest and connected to the two-wire bus with snap buttons. All sensors have a stainless-steel dry electrode. The diameter of each electrode in contact with the skin is 28 mm. To start a recording, the vest is simply put on naked skin (neither skin preparation, nor shaving) and closed with the zipper. The sensors automatically start to measure the ECG leads when they get in contact with the skin. One of the sensors (the master sensor) collects the ECG signals and sends them via Bluetooth to an external unit (e.g., a computer or a tablet), where the data is displayed and recorded for further analysis. As shown in Fig 1, ECG lead II is also displayed as a continuous wave below the 12-lead complexes.

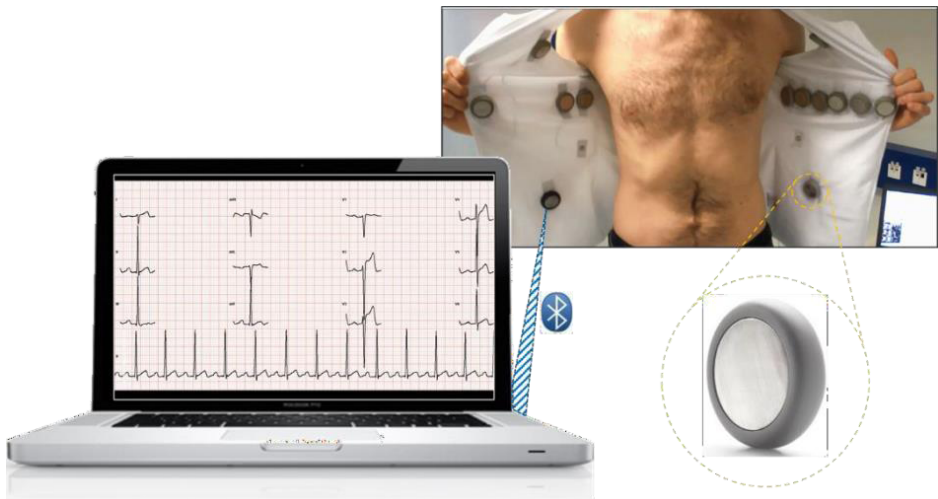


Figure 1. 12-lead ECG system under test made of ten dry-electrode cooperative sensors embedded in a tight-fitting vest [5].

The ECG-quality assessment was performed on three healthy male subjects. The clinical standard 12-lead ECG System Cardiovit CS-200 (Schiller, Baar, Switzerland) was used as a reference. The gel electrodes of the reference system were glued onto the skin at locations as close as possible to those of cooperative sensors (see measurement setup shown in Fig. 2). The subjects underwent the following three-steps protocol: 1) a few minutes resting in supine position; 2) exercise stress test on a stationary bicycle (starting at 50 W, increasing the intensity in 25 W-steps, each one 2 min long, and up to the subject's capability); 3) recovery in supine position.

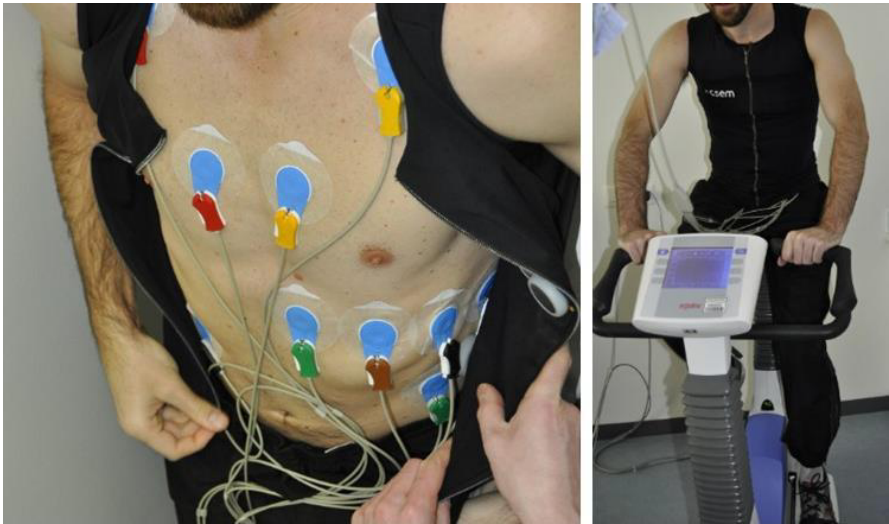


Figure 2. Test setup with reference system (left) and subject performing the ECG stress test (right).

3. Results

Fig. 3 shows excerpts of signals recorded on subject 1 with the reference system (black curves) and with the dry-electrode cooperative sensors (red curves) before (left) and during (right) the stress test. The ECG curves in Fig. 3 exemplify the general qualitative observation that the signals acquired from the two systems are very similar. As shown in the right graph in Fig. 3, similar noise appears simultaneously on both signals, which means that the electrical artefact is already present on the body. Differences in the shape of the signals (see for example Fig. 4, which shows a ventricular ectopic beat in the lead II and lead V2) are best explained with the position of the electrodes that cannot be exactly the same for the reference system and cooperative sensors.

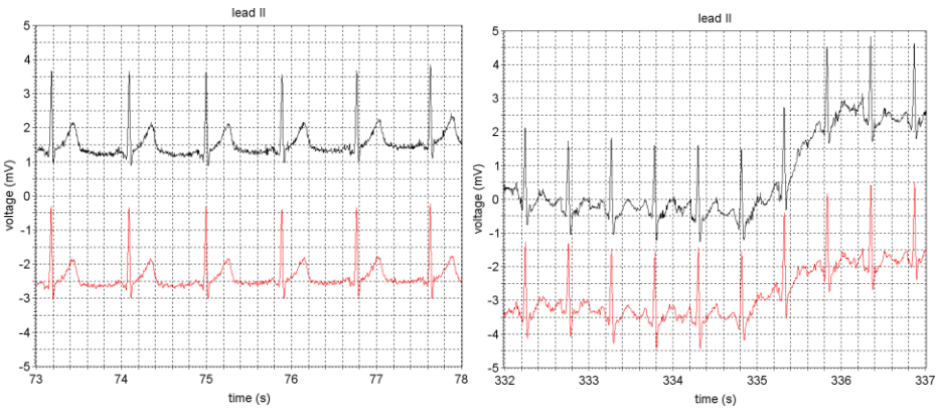


Figure 3. Typical ECG measured on the first subject with the reference system (black) and dry-electrode cooperative sensors (red): (left) supine position at rest before stress test and (right) during the stress test.

Fig. 5 shows that the baseline wandering is similar for both systems. However, the baseline of the data acquired with the Cardiovit CS-200 may be improved via signal processing with their baseline stabilizer that reduces baseline fluctuations. However, the purpose of this evaluation is not to compare the signal quality after post-processing but to qualitatively compare the ECG acquires with both systems that have an equivalent bandwidth.

Qualitative observation on Fig. 6 shows that ECG signals acquired from both systems on the second (left) and third (right) subjects during the exercise stress test are very similar. We can also notice that similar noise appears simultaneously on signals from both systems.

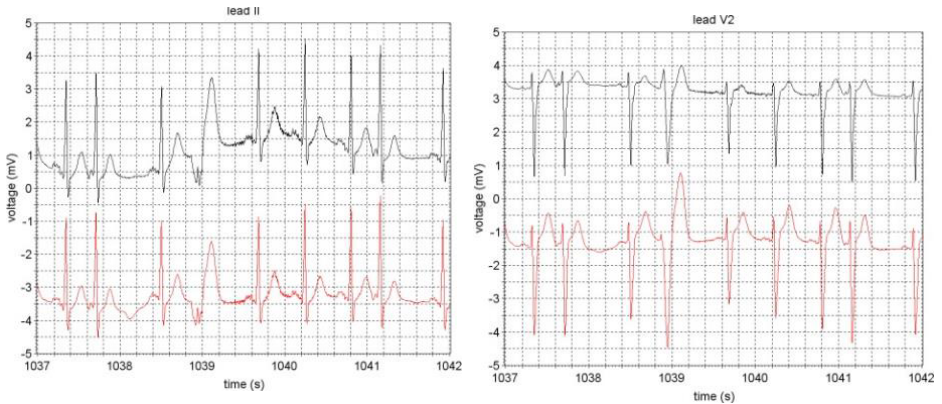


Figure 4. Typical ECG measured on the first subject with the reference system (black) and dry-electrode cooperative sensors (red): (left) ventricular ectopic beat seen from lead II and (right) same ventricular ectopic beat seen from lead V2.

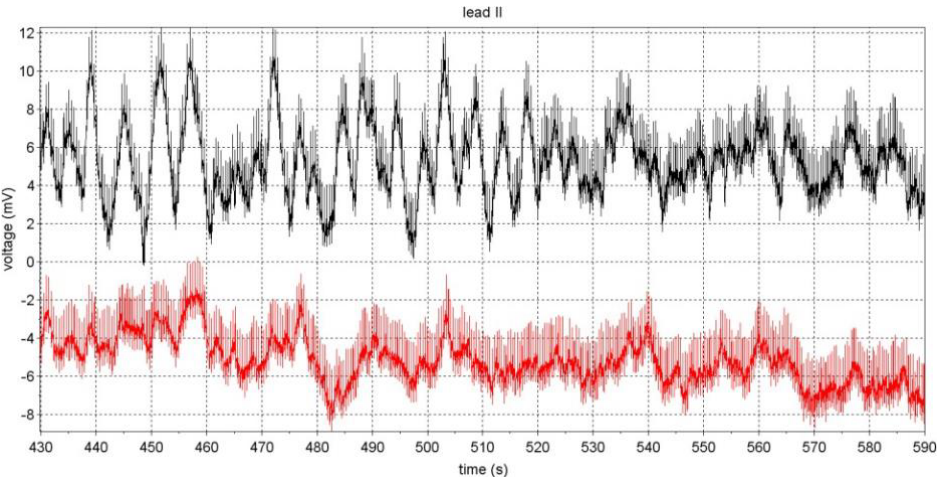


Figure 5. Comparison of baseline wandering during motion (160 s measurement on the first subject). Cooperative sensors (red) behave slightly better (no other filter than the basic 0.05 Hz to 150 Hz first-order band-pass filter) than the reference system (black).

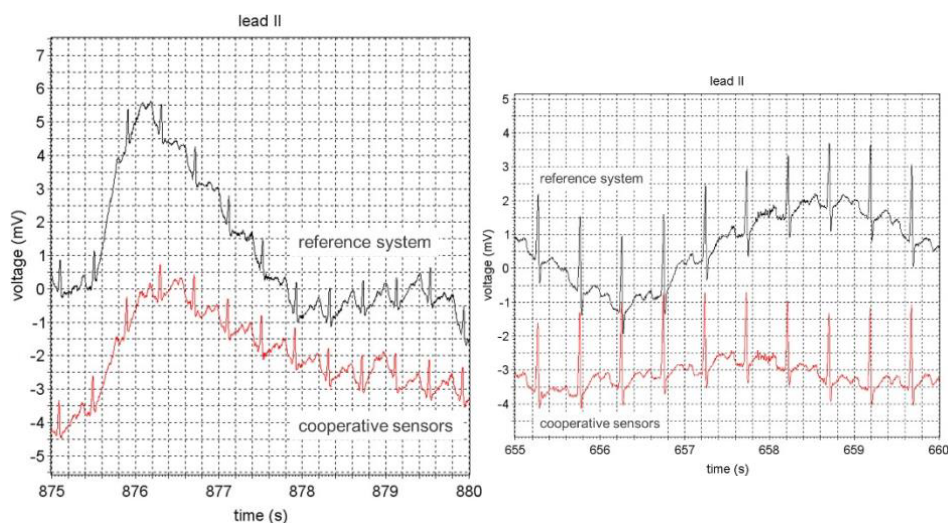


Figure 6. Typical ECG measured with the reference system (black) and dry-electrode cooperative sensors (red) on the second (left) and third (right) subjects during the exercise stress test.

4. Conclusion

This paper demonstrates the feasibility of measuring a 12-lead ECG with dry-electrode cooperative sensors in a real-life setting. The observations performed on three healthy subjects qualitatively shows that the dry-electrode cooperative-sensor technology provides signals—at rest and during exercise stress tests—of quality equivalent to conventional 12-lead ECG systems. Further studies must be carried out in order to formally validate these promising preliminary observations.

References

- [1] J. S. Steinberg, et al., ISHNE-HRS expert consensus statement on ambulatory ECG and external cardiac monitoring/telemetry, *Heart Rhythm* **14**(7) (2017), e55–e96.
- [2] N. Meziane, et al., Dry electrodes for electrocardiography, *Physiol. Meas.* **34**(9) (2013), R47–R69.
- [3] M. Rapin, et al., Two-Wire Bus Combining Full Duplex Body-Sensor Network and Multilead Biopotential Measurements, *IEEE Trans. Biomed. Eng.* **65**(1) (2018), 113–122.
- [4] T. Degen and H. Jäckel, Continuous Monitoring of Electrode–Skin Impedance Mismatch During Bioelectric Recordings, *IEEE Trans. Biomed. Eng.* **55**(6) (2008), 1711–1715.
- [5] M. Rapin, et al., Electromagnetic disturbances rejection with single skin contact in the context of ECG measurements with cooperative sensors, *EMBC conf.* (2017), 4427–4430.
- [6] R. E. Mason and I. Likar, A new system of multiple-lead exercise electrocardiography, *Am. Heart J.* **71**(2) (1966), 196–205.

Gesture Classification for a Hand Controller Device Using Neural Networks

Stelian NICOLA^{a,1}, Oana Sorina CHIRILA^a and Lacramioara STOICU-TIVADAR^a

^a*Department of Automation and Applied Informatics, University of Politehnica Timisoara, Romania*

Abstract. This paper presents a convolutional neural network-based classification of the hand flexion and extension gestures used in wrist recovery after injury. The hand gesture recognition device used in our study is the Leap Motion controller. The Leap Motion device's inability to accurately differentiate the left hand from the right hand when performing hand rotation gestures was eliminated by introducing hand and thumb direction vectors into the database used to train the neural network. A 3D environment was created for the introduction of the data describing the gesture into the database. A classification accuracy of 95% was achieved for the hand flexion and extension gesture divided into three levels for each hand. The populated database may also be used to classify other gestures involving hand rotation.

Keywords. Gesture, neural network, leap motion, hand recovery

1. Introduction

As substantiated by the literature, several sensors are used to detect hand gestures. A Flexible Epidermal Tactile Sensor is used to detect five gestures: finger flexion and extension, hand flexion and extension, and the non-state gesture (a gesture that expresses no particular state of the hand) [1]. To classify such gestures, a neural network consisting of 10 input layers was built for each gesture and an arbitrary number of hidden nodes was chosen. A set of 450 data was used to train the neural network, and 50 data entries were used for testing. A gesture classification accuracy of 88% [1] was obtained for the neural network. Image processing is also used to recognize hand gestures. In [2] a neural network is used to detect gestures representing letters, numbers or punctuation marks. Three thousand nine hundred (3,900) images were used to train the neural network. A gesture classification accuracy of 90% was achieved. Numbers and letters of the alphabet expressed by dynamic gestures may also be recognized by using the Microsoft Kinect device. The use of a radial basis function (RBF)-based neural network to classify gestures representing numbers or letters is described in [3]. The experimental results led to an accuracy of 97.25% for the gestures representing the English language alphabet, respectively of 92.63% for the gestures representing the Arabic numerals. The Microsoft Kinect device may be used in hand recovery –hand deviation exercise [4] as well as to monitor the users' correct sitting posture at a computer [5].

¹ Corresponding Author, Stelian NICOLA, University Politehnica Timisoara, P-ta Victoriei no. 2, Timisoara, Romania; E-mail: stelian.nicola@aut.upt.ro.

Judging by the degree of involvement of the limbs and by how recovery is performed, the devices used in recovery may be classified into four categories: passive, active, haptic and coaching devices [6]. Coaching devices are the cutting-edge technology in recovery. They are based on the use of cutting-edge technologies such as virtual reality, augmented reality or mixed reality to train patients on how to perform the recovery exercise. Moreover, they provide feedback to the users on the accuracy of the exercise being performed. The Leap Motion device may also be assigned to this category. It may be used in hand recovery inasmuch as it recognizes hand gestures, but also in monitoring or classifying the evolution of a disease [7] (e.g. classification of Parkinson disease). New recognizable gestures may be implemented for it [8]. The software of this device has four predefined gestures: *Screen_Tap*, *Key_Tap*, *Swipe* and *Circle*. One downside to it is that when using gestures that involve hand rotation, it confuses the left and the right hand. One immediate solution for this issue is the use of 2 Leap Motion devices [9]. The 2 devices are placed on a stand so as to be perpendicular to one another (the first is placed horizontally and the second vertically). The classification of the gestures implemented for the Leap Motion device may be performed with the aid of neural networks [10].

The objective of this paper is to achieve a specific classification for the hand flexion and extension gesture detected by means of a neural network. The Leap Motion device's main issue of recognizing the left hand from the right hand will be eliminated by entering the data corresponding to the hand and thumb rotation direction in the database. The neural network will classify the recovery level 1, 2 or 3 for each hand. The major advantage of the implemented neural network is that it is able to accurately identify the hand being used. Another benefit lies in building the database, which may also be used in training other neural networks for the classification of other types of gestures involving hand rotation.

2. Method

The accuracy of the neural network depends on: the selection of the relevant data describing the flexion and extension gesture and the number thereof, the computation of the number of neurons for the hidden layers of the neural network as well as on the division of the gesture into levels so as to make it usable/functional/useful in recovery. We explain below how the neural network was parameterized in order to achieve the best possible classification accuracy.

2.1. Neural Network

For the clusterization of the hand flexion and extension gesture by levels, we used an open-source machine learning platform (*Tensorflow*), which was parameterized. The neural network consists of 13 input layers and six output layers (Figure 1). A variant on two levels including each 17 neurons was chosen for the hidden layers. The number of neurons for the hidden layers was determined by the following formula:

$$N_h = \frac{N_s}{\alpha(N_i + N_o)} \quad (1)$$

where: N_h – number of hidden neurons;

N_s – number of data in the database (1737);

N_i —number of inputs (13)

N_o —number of outputs (6)

α —an arbitrarily chosen number, taken from the 2-10 range, in our case α is 3 (selected by former experience).

Training and testing the network required the use of the parameterized open-source Tensorflow library. A hand flexion and extension gesture detection accuracy (level classification) of 95% was achieved. A set of 1737 data was used to train the neural network and 786 data entries were used to test it.

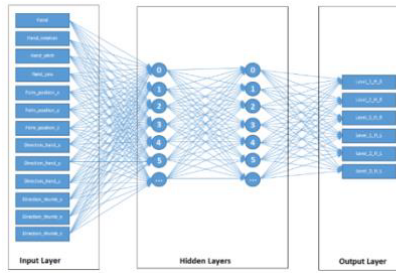


Figure 1. The Neural Network

2.2. Database and 3D environment

To devise gestures in the 3D Unity Editor for the Leap Motion device we integrated the Leap Motion version 2.3 Software Development Kit (SDK) into the editor. A 3D game was created in the Unity Editor, based on the hand flexion and extension movement, C# and PHP (Hypertext Preprocessor) scripts were written for functionality, and Python scripts to train the neural network. To have the device work as correctly as possible it was calibrated to a performance of 110 frames per second.

A MySQL database containing 13 types of data describing the hand flexion and extension gesture was created. The data types contained in the database are the following: —left or right hand, hand rotation, hand pitch, and hand yaw, palm position on all the three axes, hand direction and thumb direction. The last six input data, which refer to direction, provide higher precision in hand detection —left or right hand. In terms of data describing the gesture, a choice was made to use the hand and thumb direction because it was noticed that in some cases the Leap Motion software confuses the right hand and the left hand. Thus, such confusion was eliminated and when a hand is detected upon examination of the gesture all the elements of the other hand are deleted by the C# *Destroy(object)* method. To calculate the hand direction or the thumb direction we used the following formula:

$$D_i = \frac{p_i}{\beta} \quad (2)$$

where: D_i —hand/thumb direction and i refers to the x , y or z axis

p_i —hand/thumb position and i —refers to the x , y or z axis

β —angle calculated based on the following formula: $\beta = \sqrt{x^2 + y^2 + z^2}$

To enter the gesture description data a 3D environment (Figure 2) was created with the aid of the Unity Editor, where the users have to control an aeroplane so that it passes through a series of obstacles. The 3D object is controlled by performing left- or right- hand flexion or extension movements. 2 virtual 3D hands from the Leap Motion software package that copy the movement of the real hand were used to allow the users to constantly observe how they move their hand. The Unity Editor and the database were connected via a C# script which refers to another PHP script (Hypertext Preprocessor).

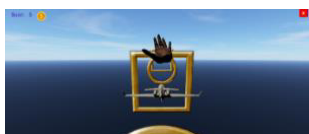


Figure 2. The 3D environment

Certain hand rotation angle ranges were selected, as shown in Table 1 to filter the data based on levels for each hand. The name assigned to each level was given in the following form: *Level_n_H*, where *n* represents level 1, 2 or 3 and *H* represents the left hand *-L* or the right hand *-R*.

Table 1. Ranges for each level of the flexion and extension gesture

Level	Range
Level_1_R	[-180; -60)
Level_2_R	[-60; 60)
Level_3_R	[180; 60]
Level_1_L	[180; 60]
Level_2_L	[60; -60)
Level_3_L	[-60; -180]

A division of the hand flexion and extension gesture into several levels was made because in hand and wrist recovery patients are initially unable to perform a full hand flexion and extension movement. Thus, recovery progress may be observed when moving from one level to the next.

The C# *RiggedHand* script was rewritten to achieve best possible results in the classification of the hand flexion and extension gesture, eliminating, therefore, the possibility for the virtual hands which copy the movement of the real hands to take any virtual position within the 3D space. Such operation was made because as far as the hand rotation gestures are concerned, the data related to distances and positions of the fingers are not important. It may be said that the virtual hands are mobile, but they have a single rotation point.

3. Results and discussion

Table 2 shows the results obtained for each level of each hand. The following metrics were computed: precision, recall, f1-score. It can be noticed that the relation between the training data number is directly proportional to precision. The maximum F1-score is of 0.98 for 675 training data for level 2 of the left hand, while the minimum F1-score is of 0.47 for level 3 of the right hand. Thus, the larger the training data set, the better the F1-score.

Table 2. Precision, recall, f1-score and number of training and test data

Hand Level	precision	recall	f1-score	No. of training data	No. of test data
Level 1 –right	0.94	0.99	0.96	148	152
Level 2 –right	0.99	0.92	0.95	473	146
Level 3 –right	0.31	0.92	0.47	201	12
Level 1 –left	0.83	0.58	0.68	127	52
Level 2 –left	0.99	0.98	0.98	675	368
Level 3 –left	0.92	0.96	0.94	113	56

This paper focused on the creation of a neural network for the classification of the hand flexion and extension gesture recognized by the Leap Motion device. A gesture classification accuracy of 95% was achieved. Emphasis was placed on the correct detection of the hand being used to perform the gesture, eliminating the confusion between the right and the left hand. The database created to train the neural network allows the users to also train other neural networks in order to classify other types of hand rotation gestures (e.g. pronation and supination). The gesture was divided into several levels so as to enable using the gesture in patients' hand and wrist recovery. Our next objective is to use machine learning techniques to detect multiple gestures performed with both hands for the Leap Motion device.

References

[1] B.R. Shin, H.S. Son, S.P. Lee, H.S. Han, A Gesture Recognition System Using a Flexible Epidermal Tactile Sensor Based on Artificial Neural Network, *International Conference on Robotics and Automation Sciences*, Hong Kong, 2017.

[2] Rosalina, L. Yusnita , N. Hadisukmana , R.B. Wahyu , R. Roestam, Y. Wahyu, Implementation of Real-Time Static Hand Gesture Recognition using Artificial Neural Network, *4th International Conference on Computer Applications and Information Processing Technology CAIPT*, Assoc Higher Educ Informat Comp, Indonesia, 2017.

[3] F. Liu, W Zeng, C. Yuan, Q. Wang, Y. Wang, Kinect-based hand gesture recognition using trajectory information, hand motion dynamics and neural networks, *Artificial Intelligence Review* **52** (2019), 563-583.

[4] M. Zabri Abu Bakar, R. Samad, D. Pebrianti, M. Mustafa, N.R.H. Abdullah, Computer Vision-Based Hand Deviation Exercise for Rehabilitation, *5th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2015.

[5] N. Gal-Nadasan, V. Stoicu-Tivadar, D.V. Poenaru, D. Popa-Andrei, E. Gal-Nadasan, Computer Aided Posture Screening Using the Microsoft Kinect on Professional Computer Users for Malicious Postures, *7th International Workshop on Soft Computing Applications*, 2016.

[6] A. H. A. Kareem, A. S. Adila, G. Husi, Recent Trends in Robotic Systems for Upper-Limb Stroke Recovery: A Low-Cost Hand and Wrist Rehabilitation Device, *2nd International Symposium on Small-Scale Intelligent Manufacturing Systems (SIMS)*, Ireland, 2018.

[7] A. H. Butt, E. Rovini, C. Dolciotti, G. De Petris, P. Bongioanni, M. C. Carbonecini and F. Cavallo, Objective and automatic classification of Parkinson disease with Leap Motion controller, *BioMedical Engineering OnLine*.

[8] S. Nicola, O.S. Lupse, L. Stoicu-Tivadar, Novel Gesture Interaction Using Leap Motion in 3D Applications, *12th IEEE International Symposium on Applied Computational Intelligence and Informatics*, 2018.

[9] M. Polsinelli, P. A. Banchetti, A. Cacchio, V. Calvisi, C. Marini, et al., Hand movement parameters calculated by the LEAP based Virtual Glove, *International Symposium on Medical Measurements and Applications (MeMeA)*, 2018.

[10] S. Nicola, O.S. Lupse, L. Stoicu-Tivadar, Enhancing Precision in Gesture Detection for Hand Recovery After Injury Using Leap Motion and Machine Learning, *Studies in health technology and informatics*, **262** (2019), 320-323.

Natural Language Processing for Detecting Medication-Related Notes in Heart Failure Telehealth Patients

Alphons EGGERTH^{a,b,1}, Karl KREINER^a, Dieter HAYN^a, Bernhard PFEIFER^{a,c},
Gerhard PÖLZL^d, Tim EGELSEER-BRÜNDL^c and Günter SCHREIER^{a,b}

^aAIT Austrian Institute of Technology GmbH, Graz, Austria

^bGraz University of Technology, Graz, Austria

^cLandesinstitut für Integrierte Versorgung – LIV Tirol, Innsbruck, Austria

^dDepartment of Internal Medicine III, Cardiology and Angiology, Medical University
Innsbruck, Innsbruck, Austria

Abstract. Heart Failure is a severe chronic disease of the heart. Telehealth networks implement closed-loop healthcare paradigms for optimal treatment of the patients. For comprehensive documentation of medication treatment, health professionals create free text collaboration notes in addition to structured information. To make this valuable source of information available for adherence analyses, we developed classifiers for automated categorization of notes based on natural language processing, which allows filtering of relevant entries to spare data analysts from tedious manual screening. Furthermore, we identified potential improvements of the queries for structured treatment documentation. For 3,952 notes, the majority of the manually annotated category tags was medication-related. The highest F1-measure of our developed classifiers was 0.90. We conclude that our approach is a valuable tool to support adherence research based on datasets containing free-text entries.

Keywords. Adherence, heart failure, telemedicine, text mining, machine learning.

1. Introduction

Heart Failure (HF) is a severe chronic disease of the heart. For affected patients, an active change of lifestyle regarding nutrition and physical activity as well as individually tailored medication is recommended in the Treatment Guidelines published by the European Society of Cardiology [1]. Telehealth networks like the HF disease management program *HerzMobil Tirol* support patients and all stakeholders involved in their treatment by implementing a closed-loop healthcare paradigm [2]. In a recent publication, Köhler et al. provided evidence that such a telehealth system can improve the outcome of HF treatment [3].

The main aims of HF telehealth solutions typically relate to a) outcomes and b) economics-related targets. These targets can be achieved by early detection of acute HF events. Besides vital sign monitoring, documentation of patients' medication and their

¹ Corresponding Author: Alphons Eggerth, AIT Austrian Institute of Technology GmbH, Reininghausstraße 13, 8020 Graz, Austria, E-Mail: alphons.eggerth@ait.ac.at.

adherence to it are major features to achieve this goal. Medication management is a crucial part of each HF telehealth system. However, medication regimes are often complex, as they are subject to drug-drug-interactions, and should be tailored to vital parameters, individual circumstances, co-morbidities, etc.

Telehealth data may comprise structured data (e.g. vital signs) as well as unstructured, textual entries (e.g. free text notes). For efficient monitoring of drug adherence, data on medication need to be available in a structured fashion. Therefore, *HerzMobil Tirol* is linked to the Austrian drug database (SISIX – Spezialitäten-Informations-System, Apotheker-Verlag) and regimes of drug intake can be specified by the physician. Patients, on the other hand, are requested to document the correct intake of drugs. However, a relevant part of the information regarding medication and adherence is gathered via phone calls by involved nurses, who subsequently create free text notes within the *HerzMobil Tirol* database, mostly, because of difficulties to provide complex information from the patients' side on drug compatibility and adherence as well as from the stakeholders' side on possible changes in the procedure and medication within the structured variables as provided by the telehealth systems. Screening of free-text entries can be used to identify missing input fields, which should be added to the queries for structured treatment documentation. To avoid tedious manual screening of numerous free texts, natural language processing and machine learning can be applied.

In a recent systematic review, Sheikhalishahi et al. give an overview of natural language processing of clinical notes on chronic diseases [4]. The biggest fraction of papers resulting from their literature search was related to diseases of the circulatory system. In the context of HF, various objectives were addressed through natural language processing. Garvin et al. [5] measured the quality of HF inpatient care based on various document types. Watson et al. [6] used electronic health record data to identify psychosocial risk factors related to a higher hospital readmission risk. While those papers deal predominately with in-clinic data, studies have been performed in our group, which focus on the specific requirements in HF treatment via telehealth. Modre-Osprian et al. [7] analysed collaboration notes of stakeholders in their HF treatment network and identified medical and organisational issues. Gruber et al. [8] designed a pipeline to classify free-text notes into task and status categories.

In this study, we analysed the content of all free-text notes found in a given dataset of *HerzMobil Tirol* with respect to medication and adherence information. We tagged the notes with pre-defined categories to potentially reveal missing input fields of the queries for structured treatment documentation. Furthermore, based on the first level of tagging, we developed a classification algorithm, which can be used in subsequent studies to automatically filter numerous free texts to a smaller subset of relevant entries.

2. Methods

2.1. Dataset

De-identified (as described in [8]) data from phases I-IV of the Austrian HF disease management network *HerzMobil Tirol* were used. *HerzMobil Tirol* connects several stakeholders who are involved in the treatment of patients (see Figure 1).

The *HerzMobil Tirol* disease management network started in 2012 with project phase I (feasibility), followed by phase II (proof-of-concept), III (specification of

routine care), IV (preparation of translation to routine care) and V (translation to routine care). Currently, *HerzMobil Tirol* is part of routine care for HF patients in Tyrol.

Within *HerzMobil Tirol*, patients record vital data (blood pressure, heart rate, body weight), subjective wellbeing and medication on a daily basis, using a specific smartphone App. Data are transmitted to a backend service, where authorized health professionals can access the data. Various roles (patient, coordinator, clinician, nurse, general practitioner, relatives, helpdesk) can communicate with each other via collaboration notes. Details concerning *HerzMobil Tirol* can be found in [10].

Our dataset comprised 3,952 notes, which were created during the treatment of 106 patients, who participated in the initial four project phases (2012 – 2016) of *HerzMobil Tirol* (see Table 1).



Figure 1. The *HerzMobil Tirol* heart failure disease management network involves several stakeholders. Clockwise: coordinator, clinician, nurse, general practitioner, relatives, helpdesk. [9]

Table 1. Patient statistics for the initial four project phases of the *HerzMobil Tirol* disease management network. *One patient of project phase II participated also in project phase IV.

Project phase	Monitoring time	Patients	Age (M \pm SD)	Male / female	Notes
-	months	-	years	-	-
I	12	10	72.8 \pm 8.7	2/8	142
II	6	40	75.4 \pm 13.0	15/25	1,746
III	3	15	70.7 \pm 10.0	7/8	494
IV	3	42	66.9 \pm 11.3	8/34	1,570
Total	-	106*	71.1 \pm 12.1	32/74*	3,952

2.2. Note classification via manual annotation and natural language processing

For note annotation, seven categories were defined: *absence of the patient*, *adaptation of the thresholds for vital parameters*, *death*, *medication*, *period of in-clinic stay*, *technical problems* and *therapeutic advice*. Due to the high number of medication-related tags and since the focus of our study was on medication adherence, category *medication* was divided into five sub-categories: *dose adaptation*, *drug added to prescription*, *drug removed from prescription*, *intake-related information* and *reason for prescription change*. For each category, the notes were manually screened for relevant information and tagged correspondingly. The annotation was done with a web application called *Twister*, which has been developed with the Python-based web framework *Django* [11] and is a part of our *Predictive Analytics Toolset for Healthcare (PATH)* [12].

Automated classification was done based on the machine-learning framework *scikit-learn* [13]. A bag-of-words model was used for classification. Stop words were removed and numbers, medication and references to time and units were abstracted. Tokenization was done with *spaCy* (ExplosionAI GmbH, Berlin, Germany) based on a pre-trained German corpus. Classification was done based on linear classifiers using stochastic gradient descent learning. For each category, a binary classifier was trained.

3. Results

Table 2 shows the number of tags for each category as well as the performance measures of the classifiers developed using a total sample size of 2,049 notes (= subset of notes with at least one tag) that were split to training (80%) and test (20%) set for evaluation. While death-related notes were very rare, there were many notes on a period of in-clinic stay. The majority of tags, however, were in the *medication* category. The classifier of the *adaptation of the thresholds for vital parameters* category showed the highest F1-measure, followed by the classifiers of the *period of in-clinic stay* and *medication* categories. For the *death* category, too few samples were available for classifier training.

Table 3 shows the results of the manual sub-categorization of 956 medication-related notes. While *dose adaptations* were the most frequent sub-category, the smallest number of tags was in the *drug removed from prescription* sub-category.

An example of a medication-related note by a general practitioner is: “I adapted the medication regime because of asymptomatic hypotension.”, which was further tagged with the *dose adaptation* and *reason of prescription change* categories.

Table 2. Category tags and performance measures of the trained classifiers.

Category	Tags	Precision	Recall	F1-Measure
Adaptation of the thresholds for vital parameters	146	0.99	0.82	0.90
Period of in-clinic stay	640	0.97	0.77	0.86
Medication	956	0.91	0.80	0.85
Technical problems	354	0.93	0.75	0.83
Absence of the patient	178	0.90	0.58	0.71
Therapeutic advice	180	0.75	0.44	0.55
Death	11	-	-	-

Table 3. Results of the manual sub-categorization of the *medication* category.

Sub-category	Tags
Dose adaptation	287
Intake-related information	253
Reason of prescription change	204
Drug added to prescription	124
Drug removed from prescription	91

4. Discussion

Annotation of the free-text notes revealed that information is mainly provided regarding the *medication*, *period of in-clinic stay* and *technical problems* categories. Therefore, free text notes are a valuable source of additional information for the evaluation of medication adherence in a telehealth HF disease management network. However, manual analysis of every single note is a time-consuming task. Our automated classifiers achieved F1-measures of up to 0.90 with $F1 > 0.80$ for four out of 7 categories. Therefore, our classifiers could be used as an accurate filter of subsets containing only relevant notes, thus sparing researches from tedious screening work.

In the *medication* category, many notes contained intake-related information and revealed the reason for a prescription change. Although it is possible to document addition or removal of a drug or to change the medication dose in a structured way in the system, we found that additional information is often documented in free-text notes,

such as the reason for prescription, or complicated dosage regimes (every second day, dosage depending on patient status, etc.). Hence, a suggestion to further improve the queries for structured treatment documentation would be to add respective input fields.

Collaboration notes were manually tagged with categories by only one of the authors. To obtain a more reliable gold standard for the machine learning tasks, the notes need to be re-tagged by multiple persons independently for subsequent analyses.

In future studies, text mining algorithms should be further developed towards fully-automated information extraction from free texts, like identifying events, dates and values to automatically add them to structured data. Moreover, it needs to be investigated if doing annotation and classification on a sentence level will result in better performance.

In conclusion, our approach for automatic notes categorization is a valuable tool to support adherence research based on datasets containing free text entries, as it allows data scientists to filter a subset of relevant entries.

References

- [1] P. Ponikowski, et al., 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC (in eng), *Eur Heart J* **37(27)** (2016), 2129-2200. doi: 10.1093/eurheartj/ehw128.
- [2] R. Modre-Osprian, G. Pölzl, A. Von Der Heidt, and P. Kastner, Closed-loop healthcare monitoring in a collaborative heart failure network (in eng), *Stud Health Technol Inform* **198** (2014), 17-24.
- [3] F. Koehler, et al., Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial (in eng), *Lancet* **392(10152)** (2018), 1047-1057. doi: 10.1016/S0140-6736(18)31880-4.
- [4] S. Sheikhalishahi, R. Miotto, J. T. Dudley, A. Lavelli, F. Rinaldi, and V. Osmani, Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review (in eng), *JMIR Med Inform* **7(2)** (2019), e12239. doi: 10.2196/12239.
- [5] J. H. Garvin, et al., Automating Quality Measures for Heart Failure Using Natural Language Processing: A Descriptive Study in the Department of Veterans Affairs (in eng), *JMIR Med Inform* **6(1)** (2018), e5. doi: 10.2196/medinform.9150.
- [6] A. J. Watson et al., Linking electronic health record-extracted psychosocial data in real-time to risk of readmission for heart failure (in eng), *Psychosomatics* **52(4)** (2011), 319-27. doi: 10.1016/j.psych.2011.02.007.
- [7] R. Modre-Osprian, K. Gruber, K. Kreiner, G. Schreier, G. Poelzl, and P. Kastner, Textual analysis of collaboration notes of the telemedical heart failure network HerzMobil Tirol (in eng), *Stud Health Technol Inform* **212** (2015), 57-64.
- [8] K. Gruber, R. Modre-Osprian, K. Kreiner, P. Kastner, and G. Schreier, Development of text mining-based classification of written communication within a telemedical collaborative network (in eng), *Stud Health Technol Inform* **212** (2015), 35-42.
- [9] A. Eggerth, D. Hayn, and G. Schreier, Medication management needs information and communications technology-based approaches, including telehealth and artificial intelligence (in eng), *Br J Clin Pharmacol* (2019). doi: 10.1111/bcp.14045.
- [10] A. Von der Heidt et al., HerzMobil Tirol network: rationale for and design of a collaborative heart failure disease management program in Austria (in eng), *Wien Klin Wochenschr* **126(21-22)** (2014), 734-41. doi: 10.1007/s00508-014-0665-7.
- [11] K. Kreiner, D. Hayn, and G. Schreier, Twister: A Tool for Reducing Screening Time in Systematic Literature Reviews (in eng), *Stud Health Technol Inform*, **255** (2018), 5-9.
- [12] D. Hayn et al., Predictive analytics for data driven decision support in health and care, *it-Information Technology* **60(4)** (2018), 183-194.

Patient-Generated Health Data in the Clinic

Nirusha JEEVANANDAN^{a,1} and Christian NØHR^a

^aThe Maersk Mc-Kinney Møller Institute, University of Southern Denmark, Denmark

Abstract. Physician appraisal of Patient-Generated Health Data (PGHD) is unclear. It is not known how willing they are to consider self-registered measurements prior to treatment. In this research, a semi-structured interview was designed and administered to 10 physicians in Denmark. The study found that physicians are willing to work with unsolicited PGHD and integrate it as a part of the treatment if evidence proved the data to be valid and reliable. They request documented evidence for the use of each PGHD device and mobile application.

Keywords. Patient-generated health data, eHealth, Medical wearables

1. Introduction

The health care system is awaiting a revolutionizing paradigm shift, where it becomes more personalized and enable citizens to take more responsibility for their health and well-being. Digitalization in the health sector is increasing all of Europe as personalized medicine, digital innovations, the Internet of things, and artificial intelligence gains momentum [1]. This shift transfers the control of health care from hospitals and physicians into the hands of patients; more specifically, into their homes and mobile devices [2]. It is estimated that the global market for medical wearables will reach \$19.5 billion in 2021 [3]. Patients are becoming more self-reliant in using digital opportunities to maintain their health [4]. Patient-Generated Health Data (PGHD) can be obtained digitally using various devices and mHealth applications. PGHD are health-related data created, recorded, or gathered by or from patients [5]. PGHD is often confused with patient-reported data (PRO-data). In Denmark, PRO-data is described as data about the patient's health, including physical and mental health, symptoms, health-related quality of life, level of functioning, as reported directly by the patient. The collection and reporting of data are requested by the health practitioners. In contrast, PGHD is initiated and driven by individuals, hence a patient-driven data collection modality [6].

It is assumed that PGHD can be instrumental in changing the health care system to focus on health care rather than sick care and to assist patients in being able to prevent many health problems before they become manifest [7]. Despite the growth in the use of PGHD, a literature search indicated a knowledge gap and a lack of evidence-based literature examining physicians' appraisal of PGHD in Denmark. However, the use of PGHD is more common in other countries, mostly in the US, where several health care professionals have worked with PGHD for clinical usage [8]. By accepting and utilizing PGHD, it is possible for physicians to gain greater insight into patients' overall health,

¹ Corresponding Author, E-mail: nirusha.jeevan@gmail.com

from which the physicians can make more informed decisions [9]. Examples of PGHD can be physical activity, sleep, temperature, heart rate, electrocardiogram (ECG), blood pressure, blood glucose, and oxygen saturation. In addition to data captured by sensors, PGHD can be supplemented by patient entered information such as mood [10]. More active patients will result in an increasing number of appointments with self-recorded measurements [7]. The benefits of PGHD can be limited for patients if their physicians do not agree to utilize the collected data, and physicians' appraisal of PGHD is still unclear in many countries. It is not known how willing they will be to consider these measurements prior to treatment, hence the aim of this study is to examine physician's appraisals of unsolicited PGHD and investigate how these data can be applied in the clinic in collaboration with patients.

2. Methods and material

Since the aim of this research is to describe and understand clinicians' subjective appraisal of practical PGHD use, qualitative interviews were chosen as the method. An interview invitation was sent to 17 possible participants from the author's professional network in Denmark. Attention was paid to variation in age and field of specialization. In qualitative studies, it is not possible to determine the minimum sample size, but it was assumed that data saturation could be reached within the 17 respondents. An interview guide reflecting the research questions was developed for the semi-structured interviews. In order to test the set of interview questions, two pilot interviews were performed. Explanations for key terms for PGHD, PRO-data, mHealth, pHealth, and patient empowerment were prepared and used for unanimously describing the terms during the interviews. The need for this description indicates how known these terms were.

10 of the 17 invited agreed to participate, and interviews were conducted in the period from 21-03-2019 to 10-04-2019. The physicians decided the date and place for the interviews. All participants provided consent to the interview and audio recording. The interviews ranged in length from 7 to 36 minutes. The interviews were transcribed verbatim for later analysis.

A deductive analysis approach was applied to derive information on categories from the interview guide, priori concepts, and pre-existing knowledge. The themes and coding categories were: 1) current knowledge of PGHD, 2) Physicians' perspective of PGHD, 3) Future use of PGHD.

To guide the questions and have a specific focus for the interviews, the ECG measurement facility on the Apple Watch, series 4, was demonstrated and used for reference during the interviews. Additionally, pictures of the ECG reading were shown, in the health application and in a PDF file.

3. Results

Generally, the physicians were open-minded about PGHD and the increasing use of mHealth and growing patient empowerment. Some of the perceived benefits mentioned are patients being more liberal, active in maintaining their health, can result in less overload for the physician and also be an assisting tool for the physician. One of the physicians mentions the following: *"I actually think there are some things that might be*

really great if the patients could monitor themselves because then they do not have to show up here to find out things and I also believe in that way it is the relevant patients which come to you, it is not just everyone that comes in, and then you think I am sorry, but I cannot really help you... Whereas you can be some kind of gatekeeper from the start and then the relevant patients come in, now I have this, you can see this is what I have, I want to come in and talk to you."

The results of the analysis of the interviews are presented in Table 1. Some of the concerns expressed were that it could create unnecessary worrying among patients and over-observations, concerns about an increase in the inequality in health and about the

Table 1. The different aspects investigated in the study

Item	Aspect/construct	Responses (n=10)	
1	Not familiar with PGHD	%	70
2	Confused PGHD with PRO	%	30
3	Familiar with the existence of PGHD	%	10
4	Experiences of patients seeking out to them with health-related data generated by themselves without prior being solicited from the physicians	%	40
5	Experienced the increasing use of mHealth and patient empowerment	%	70
6	Worried about patients over-diagnosing themselves and creating unnecessary	%	70
7	Worried about increase/creating an inequality in health	%	40
8	Feels it is a good idea that patients can register own data when symptoms occur	%	20
9	Concerned about data credibility and needs validation of data	%	80
10	Willing to work with PGHD in the future as a part of patient treatment if they find data	%	80

data validity. The majority of the physicians do not trust data validity for the Apple Watch's ECG application and indicated that they would perform their ECG measurements.

The physicians have different opinions about the future use of PGHD. Most of them are willing to use it but have different criteria and necessities. Additionally, physicians reported different ideas for future use where PGHD is included as part of treatment, as well as perspectives on how the data should be integrated as a part of the EHR.

4. Discussion

Overall the physicians were not familiar with the term PGHD. Those who said they were actually confused it with PRO-data, which some of them were familiar with to a certain extent. Although some of the physicians were not familiar with PGHD and the use of it, they already had experienced patients bringing their self-registered data without it being solicited by their physician. Most of the physicians had experienced increasing patient empowerment, except for two of them who work with elderly and mentally ill patients. Most of the physicians were familiar and experienced with the patients' use of mobile health applications and patients looking up their symptoms and diagnosis on Google.

Generally, the physicians are open-minded about PGHD. Some of the benefits and concerns expressed in the interviews replicate results reported in the literature [6, 8], [11–13]. Benefits can be gained not only for the physicians but also for the patients, by

being more active and involved in the decision making for their treatment, thereby becoming more patient-centred. This can likewise help physicians to be diagnostic efficient and effective. The data validity and verification were one of the main concerns mentioned by the majority of the physicians, and it was decisive for their willingness to use PGHD. Concerns about over-diagnosing, patients becoming excessively worried, and making themselves feel more ill than they were, in reality were mentioned repeatedly. Even though the physicians felt that patients became more concerned from generating their data, some of the physicians still had an understanding of why they did so.

Another major concern was an increase of inequality in health among patients. The increasing inequality and deepening of the gap between the “haves” and “have-nots” are also one of the main apprehensions mentioned by Eysenbach as a considerable threat [14]. The concern of the availability of PGHD among low-income and elderly populations and those who often needed it the most was also shared among physicians’ opinions found in the literature [11]. A cardiologist at University of California San Francisco states that the individuals who purchase Apple watches often are those who are already health-conscious and probably quite healthy, whereas those who would have the most benefit from the atrial fibrillation screening tend to be the elderly, who might be less tech-savvy and less concerned about their health [15]. In addition to worried patients, this situation can also induce a greater workload for the physicians.

Many of the physicians emphasize the importance of not only looking at measurements but also at the patients’ symptoms. They would consider forwarding the patient to a cardiac specialist if the symptoms were significant and the ECG still showed a sinus rhythm. Generally, in this research, it was discovered that most of the physicians did not fully trust the ECG application on Apple Watch. Most of them were open to looking at the ECG reading brought by the patients but would then use their own verified device, e.g. a Holter. However, according to Cardiologist John P. Higgins, the updated capabilities of the Apple Watch Series 4 can broaden the monitoring capabilities of existing technologies. He states that *it is less invasive than Holter monitor, has great battery life and can be with the person 24/7* [16].

American physicians share the concerns of their Danish colleagues about the lack of evidence to justify the Apple Watch ECG against the risk of false positives. The latter would cause healthy/wealthy people to take up the time resources of both physicians and hospitals since there is no evidence that the benefits outweigh the harm. At the same time, another American physician states that for people with heart conditions or symptoms, the Apple Watch could serve as a good tool for detecting when something is wrong and when a visit at the physician is necessary [15]. This research shows that the physicians are willing to use PGHD as a part of treatment, as long as their individual criteria and needs are met. Some of the needs reported in the literature [13], such as making time to review PGHD, uncertainty of data quality and the need to distinguish between data measured by patients and by health care professionals.

Most physicians agreed with integrating PGHD into the EHR, as was already done according to [17]. Many of their future ideas for the use of PGHD is likewise based on importing the data into the EHR. However, it was important for physicians to be able to distinguish between the data collected by themselves and by the patients.

Despite the different opinions about PGHD and the benefits that can be gained by using it, and likewise the increased workload and burden for the physicians, it is clear that the first priority among the physicians is the safety of the patients. They do not wish to place all responsibility on the patients. They want to make sure that the devices and mobile applications can be trusted while avoiding unnecessary patients worry.

5. Conclusion

This qualitative study showed that Danish physicians sampled have an overall positive appraisal of PGHD for clinical usage. Despite the many clinical benefits to be gained, the physicians prioritize data from validated, and reliability checked sources before taking it into consideration for clinical decisions. The physicians appreciate increasing patient empowerment and are willing to use PGHD as a part of the treatment for those who have the resources and desire to do so. Additionally, the physicians welcome importing PGHD into the EHR, as long as they are clearly marked. The physicians request more studies and documented evidence for the use of each PGHD device and mobile application; therefore, further research is needed to achieve a general appraisal.

References

- [1] *Healthcare 4.0 and the future of life sciences*, <https://www.aspeninstitute.it/en/programs/healthcare-40-and-future-life-sciences.>, Aspen Institute Italia, 2017.
- [2] S. Rab, *The Paradigm Shift in Medicine and How it Affects Technology*, CIOReview, <https://healthcare.cioreview.com/cioviewpoint/the-paradigm-shift-in-medicine-and-how-it-affects-technology-nid-3823-cid-31.html>
- [3] R. Abdolkhani, A. Borda, and K. Gray, Quality Management of Patient Generated Health Data in Remote Patient Monitoring using Medical Wearables - A Systematic Review, *Stud. Heal. Technol. Informatics* **252** (2018), 1–7.
- [4] Sundheds- og ældreministeriet, Finansministeriet, Danske Regioner, and KL, *Ét sikkert og sammenhængende sundhedsnetværk for alle: Strategi for digital sundhed 2018-2022*, 2018.
- [5] *Health IT in Health Care Settings: Patient-Generated Health Data*, <https://www.healthit.gov/topic/scientific-initiatives/patient-generated-health-data.>, HealthIT, US Health and Human Services, 2018.
- [6] D. J. Cohen, S. R. Keller, G. R. Hayes, D. A. Dorr, J. S. Ash, and D. F. Sittig, Integrating Patient-Generated Health Data Into Clinical Care Settings or Clinical Decision-Making: Lessons Learned From Project HealthDesign, *JMIR Hum. Factors* **3** (2016), e26.
- [7] P. Hesseldahl, Sundhedsrevolutionen starter på din smartphone, *Mandag Morgen*, 2018.
- [8] J. Adler-Milstein and P. Nong, Early experiences with patient generated health data: health system and patient perspectives, *J. Am. Med. Informatics Assoc* (2019), 1–8.
- [9] C. Caraballo, *Patient Generated Health Data: Addressing Concerns and Moving Forward*, <https://www.himss.org/news/patient-generated-health-data-addressing-concerns-and-moving-forward?ItemNumber=47889>, Healthcare Information and Management Systems Society (HIMSS), 2016.
- [10] H. S. Murthy and W. A. Wood, The Value of Patient Reported Outcomes and Other Patient-Generated Health Data in Clinical Hematology, *Curr. Hematol. Malig. Rep* **10** (2015), 213–224.
- [11] S. Nundy, C.-Y. E. Lu, P. Hogan, A. Mishra, and M. E. Peek, Using Patient-Generated Health Data From Mobile Technologies for Diabetes Self-Management Support: Provider Perspectives From an Academic Medical Center, *J. Diabetes Sci. Technol.* **8** (2014), 74–82.
- [12] B. Holtz, K. Vasold, S. Cotten, M. Mackert, and M. Zhang, Health Care Provider Perceptions of Consumer-Grade Devices and Apps for Tracking Health: A Pilot Study, *JMIR mHealth uHealth* **7** (2019), e9929.
- [13] M. J. Reading and J. A. Merrill, Converging and diverging needs between patients and providers who are collecting and using patient-generated health data: An integrative review, *J. Am. Med. Informatics Assoc.* **25** (2018), 759–771.
- [14] G. Eysenbach, What is e-health?, *J. Med. Internet Res.* **3** (2001), 1–5.
- [15] D. Robitzski, *Doctors: Apple Watch EKG Users Could Burden the Healthcare System*, <https://futurism.com/doctor-apple-watch-ekg-healthcare>, Futurism, 2019.
- [16] *Cardiologists react to new Apple Watch ECG capabilities*, <https://www.healio.com/cardiology/arrhythmia-disorders/news/online/%7B60ccdb24-e448-4832-bcbc-f696acb31b19%7D/cardiologists-react-to-new-apple-watch-ecg-capabilities.>, Healio, 2018.
- [17] C. Cetrangol, N. Genes, Y.-F. Y. Chan, E. E. Schadt, S. Violante, and L. Rogers, From smartphone to EHR: a case report on integrating patient-generated health data, *npj Digit. Med.* **1** (2018).

Register-Based Research of Adverse Events Revealing Incomplete Records Threatening Patient Safety

Ulla-Mari KINNUNEN^{a,1}, Eija KIVEKÄS^a, Sari PALOJOKI^b, Kaija SARANTO^a

^a *University of Eastern Finland, Department of Health and Social Management, Finland*

^b *Ministry of Social Affairs and Health, Finland*

Abstract. Inadequate, missing or incorrect patient information is usually related to poor documentation. It has several negative effects on patient care processes, and, thus to quality of care, care continuity, and patient safety. It is one of the causes of patient claims. The aim of this study was to analyze patient safety reports and to find out which documentation hazards are damaging to patient safety. Data consisted of the patient incident reports (n=82 353) from seven health and social care areas from 2007–2016 in Finland. A descriptive analysis was conducted to explore the type of service provider and incidents reporting risks in patient data management and documentation. Adverse events due to patient data management and documentation were unusual; however, 18 cases were documented where patients suffered serious harm. Nearly half of the reports resulted from inadequate, missing or incorrect information. Uniform structures, documentation, and service processes need to be developed.

Keywords: Documentation, safety management, patient harm.

1. Introduction

The transition from paper-based patient documentation systems to electronic systems has been widespread during the past years both internationally [1] and in Finland [2]. Some existing evidence shows that computerized patient data management systems [3–4] and electronic documentation systems [5] affect positively patient safety and quality of care. In addition, structuring electronic health records can have an impact, e.g., on decreasing medication errors, increasing documentation quality, and on care process efficiency, care guidelines conformation and decreasing of prescription errors [6]. However, in some cases, despite classifications [7] or electronic systems [8] in use, the quality of documentation is weak [3,4]. Thus, poor or missing documentation is among the most common causes for patient claims. Poor documentation hinders the visibility and evidence of health care professionals' competencies, compliance of standardized care procedure and clinical guidelines, quality of care, the assessment of the health care provider, and the defence against claims [9].

One of the aims of the Finnish Patient and Client Safety Strategy 2017–2021 published by the Ministry of Social Affairs and Health [10] is that by 2021, all electronic health service processes and procedures will be safe, and patients and clients protected from harmful incidents. This also includes patient data management and documentation where all the essential processes in information flow and documentation must be

¹ Ulla-Mari Kinnunen, Virranniementie 3, 78310 Varkaus, Finland; E-mail: ulla-mari.kinnunen@uef.fi. +358403553953

actualized using the same formats and structures in every health and social care organization, and in information exchange between the organizations. In addition, professionals in health and social care will have access to information systems and patient or client information that support their work and their processes [11]. A large number of incidents related to information management consist of documentation errors. Inadequate, missing or incorrect information, e.g. wrong name or birthday, might stand for some specific information loss. This might have a very important role, e.g., in the continuity of care and more, in patient safety [12].

Common data structures and structured patient data benefit broadly and in various ways, both the patient, professionals, and organizations [6, 13]. The Finnish patient safety incident reporting system, HaiPro, [14] is widely used in national health and social care. The online reporting system allows reporting of incidents for professionals, but also patients and their families. The aim of this study is to analyze incident reports and especially concentrate on finding out which documentation hazards are most damaging to patient safety.

2. Methods

Data consisted of the incident reports (n=82 353) from seven health and social care areas from the years 2007–2016: three hospital districts, health and social care of two large cities, and two social and health care districts. Reports were retrieved from the Finnish Society for Patient Safety database, which collects reports from about 200 health and social care facilities. Reports include both structured and free-text descriptions of safety events.

A descriptive analysis was conducted to explore the type of service provider (primary, health and social care) and incidents reporting risks in category *Information flow and information management related*. HaiPro [14] system has 13 incident categories (Table 1). In addition, the reporter describes the incident by narrative text clarifying the type, context, and circumstances of the incident. The consequences for patients or service providers are also assessed. Reporting incidents is confidential, voluntary, and anonymous, highlighting learning from mistakes and blame-free actions. HaiPro system implementation began in hospitals in 2007, and the social service began implementing the HaiPro system just recently. Reports of inadequate, missing or incorrect information were by inpatient wards and outpatient departments.

Table 1. Categories of HaiPro system

Categories		
1	Medication or iv fluid care, blood transfusion, or contrast agent related	8 Devices or their use related
2	Information flow and information management related	9 Hygiene related/protection against infections
3	Diagnoses related	10 Injury
4	Surgical procedures related	11 Paramedical services related
5	Invasive procedures related	12 Violence
6	Other care related	13 Deviation in radiotherapy
7	Laboratory, imaging-related	

The analyzed adverse events consisted of reports of primary healthcare, special healthcare (hospitals) and social care (i.e., elderly service). Approval of the University

of Eastern Finland Committee on Research Ethics was required. SPSS 23 software was utilized for analyzing the structured data (Armonk, NY, IBM).

3. Results

Adverse events during *Information flow and information management* (n=12 294) comprised 41.2 % (n=5070) of incident reports concerning *Patient data management and documentation* (Table 2). Of those, more detailed description of the incident reports of *Patient data management and documentation* included 40.8 % (n=2066) were the reports of inadequate, missing or incorrect information. Analyses showed that the relative proportion of adverse events was similar in primary healthcare as well as secondary and tertiary healthcare. In both sections, half of the adverse events were reports from inpatient wards. Of all data obtained, most of the reports (63,2 %) came from registered nurses and physicians 4,2 percentage.

Table 2. Information flow and documentation related adverse events

Information flow and information management related adverse events (n = 12 294)	n	%
Patient data management and documentation	5070	41,2
Care coordination	4686	38,1
Verbal communication	2538	20,7
Patient data management and documentation related adverse events (n = 5 070)	n	%
Inadequate, missing or incorrect patient information	2066	40,8
Error in personal data or contact information	576	11,4
Patient information not retrieved or printed from patient records	413	8,1
Wrong or outdated information in patient records	376	7,4
Missing referral or inadequate, missing or incorrect information in referral	329	6,5
Information input or retrieval of patient records hindered	306	6
Referral of test result documentation for wrong patient	215	4,2
Patient information documented in wrong place	201	4
Other or unknown	588	11,6

When reporting, the nurse or physician picks the incident category and the applicable harm that he or she thinks has occurred to the patient. In most cases (59,2 %, n = 1948), as a consequence, no harm was done to patients due to adverse events of *Patient data management and documentation*. Patients suffered from moderate harm in 178 (5,4 %) cases and serious harm in 18 (0,5 %) cases. Five of those serious cases were related to inadequate, missing or incorrect information. Not knowing the degree of harm were 10.8 % of the cases (n=166) which related to inadequate, missing or incorrect information.

4. Discussion

Uniform, accurate, and standardized data structures and documentation are a prerequisite to safe and good-quality care [6]. Inadequate, missing or incorrect information can be

very important information. It may be essential for patient care which has an explicit connection to patient safety [3-5,12]. There are a lot of differences in the structures of patient care documentation [13]. Over one-third of the incident reports of *Information flow and information, management* were related to *Patient data management and documentation*. Of those reports, nearly half were of inadequate, missing or incorrect information. This information deficiency can be, e.g. missing information about patient's medicine or allergy, which cannot be seen in the electronic documentation system. It can also be inadequate documentation of medicine administration, a missing patient identifier in a laboratory test, or wrong electrocardiogram for a physician to interpret.

Adverse events related to patient data management and documentation included a total of 18 cases where the patient suffered serious harm. Most of these incidents were because of inadequate, missing or incorrect information. Lack of and deficiencies in documentation likely refers to poor-quality care and missing standardized care processes [7]. Adverse events related to *Patient data management and documentation* can refer to inoperative work and operational practices, which means that professionals need to be trained simultaneously when implementing a new system [10-11].

Documentation deficiencies can also be an outcome of inoperative and unclear documentation systems [7] or lack of rules and requirements for documentation [8]. These need to be taken into account in the early stages of the electronic documentation system's acquisition and when defining the common documentation guidelines for both organizational and national levels [6, 11]. Advances, such as improved data quality and patient's legal protection, decreased misinterpretations, better possibilities for data re-use and data retrieval, and possibilities for decision support connection to patient records and its use have been presented. This also applies to the use of standardized health and social care languages [6, 13].

The purpose of HaiPro system is highlighting learning from mistakes [9, 14]. Accurate documentation is evidenced to affect patient safety, and the main criteria for high-quality healthcare are considered to be well-trained staff and treatment that works [3-5]. Incidents of *Patient data management and documentation* occurred both in primary and special healthcare. Our data illustrated that the handling process requires confirmation.

5. Limitations

Data for this study represent seven health and social care districts. However, our data included safety incidents reported by the largest hospital district in Finland, which accounts for approximately one-third of the population. Due to the voluntary nature of the HaiPro incident reporting system, not all incidents are reported. In the future, more emphasis should be put on how the systems can be made more comprehensive and more effective.

6. Conclusion

Good quality and accurate documentation **are** known to affect patient safety positively. Organizations need to plan and implement standardized documentation platforms and care processes. Additionally, we need more accurate information on adverse events and understanding of the problem for improving health and social care processes. Even

though the amount of serious harm to the patient because of inadequate, missing or incorrect information was not big in this data, this example shows the value of incident reporting systems. Patient safety must never be harmed because of poor documentation.

References

- [1] L.L. Fragidis and P. D. Chatzoglou, Implementation of a nationwide electronic health record (EHR): The international experience in 13 countries, *Int J Health Care Qual Assur* **31(2)** (2018), 116-130.
- [2] T. Vehko, S. Ruotsalainen, H. Hyppönen (eds.), *E-health and e-welfare of Finland. Checkpoint 2018*, http://www.julkari.fi/bitstream/handle/10024/138244/RAP2019_7_e-health_and_e-welfare_web_4.pdf?sequence=4&isAllowed=y, National Institute for Health and Welfare (THL), Helsinki, Finland, 2019.
- [3] A. Cheung, F.H.P. van Velden, V. Lagerburg, and N. Minderman, The organizational and clinical impact of integrating bedside equipment to an information system: A systematic literature review of patient data management system (PDMS), *Int J Med Inform* **84** (2015), 155-165.
- [4] K. Donsa, P. Beck, B. Höll, J.K. Mader, L. Schaupp, J. Plank, K.M. Neubauer, C. Baumgartner, and T.R. Pieber, Impact of errors in paper-based and computerized diabetes management with decision support for hospitalized patients with type 2 diabetes. A post-hoc analysis of a before and after study, *Int J Med Inform* **90** (2016), 58-67.
- [5] B. McCarthy, S. Fitzgerald, M. O'Shea, C. Condon, G. Hartnett-Collins, et al., Electronic nursing documentation interventions to promote or improve patient safety and quality care: a systematic review, *J Nurs Manag* (2018), doi: 10.1111/jonm.12727.
- [6] R. Vuokko, P. Mäkelä-Bengs, H. Hyppönen, M. Lindqvist, and P. Doupi, Impacts of structuring the electronic health record: Results of a systematic literature review from the perspective of secondary use of patient data, *Int J Med Inform* **97** (2017), 293-303.
- [7] D.R. DeAlmeida, V.J. Watzlaf, P. AnaniaFiroozan, O. Salguero, E. Rubinstein, et al., Evaluation of Inpatient Clinical Documentation Readiness for ICD-10-CM, *Perspect Health Inf Manag* (2014).
- [8] N. Wang, P. Yu, and D. Hailey. The quality of paper-based versus electronic nursing plan in Australian aged care homes: A documentation audit study, *Int J Med Inform* **84** (2015), 561-569.
- [9] B.A. Wilbanks, M. Geisz-Everson, and R.R. Boust, The Role of Documentation Quality in Anesthesia-Related Closed Claims. A Descriptive Qualitative Study, *Comput Inform Nurs* **34(9)** (2016), 406-412.
- [10] *Government resolution. Patient and client safety strategy 2017–2021*, http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/80352/09_2017_Potilas-%20ja-%20asiakasturvallisuuksstrategia%202017-2021_suomi.pdf, Publications of the Ministry of Social Affairs and Health, Ministry of Social Affairs and Health, 2017.
- [11] *Information to support well-being and service renewal. eHealth and eSocial Strategy 2020*. http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/74459/URN_ISBN_978-952-00-3575-4.pdf?sequence=1, Ministry of Social Affairs and Health, Finland: Edita Prima, 2015.
- [12] V. Jylhä, *Information management in health care: a model for connecting information culture and patient safety*, http://epublications.uef.fi/pub/urn_isbn_978-952-61-2384-4/urn_isbn_978-952-61-2384-4.pdf, Publications of the University of Eastern Finland, Dissertations in Social Sciences and Business Studies, (accessed August 20, 2019).
- [13] K. Saranto, U.-M. Kinnunen, E. Kivekäs, A.-M. Lappalainen, P. Liljamo, E. Rajalahti, and H. Hyppönen, Impacts of structuring nursing records: a systematic review, *Scand J Caring Sci* (2013), 1-19.
- [14] Awanic Ltd. HaiPro, <http://awanic.com/eng/>, (accessed September 15, 2019).

Teletriage Pilot Study (Strategy for Unscheduled Teleconsultations): Results, Patient Acceptance and Satisfaction

Andrés Santiago FRID^{a1}, María Florencia Grande RATTI^b, Ana PEDRETTI^b, Marina VALINOTI^b, Bernardo MARTÍNEZ^b, Janine SOMMER^a, Daniel LUNA^a, Fernando PLAZZOTTA^a

^a*Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina*

^b*Internal Medicine Research Unit, Hospital Italiano de Buenos Aires, Argentina*

Abstract. Non-urgent consultations to an Emergency Department (ED) contribute to overcrowding. Telecommunications represent a potential strategy to reduce some face-to-face consultations. Objectives: To describe characteristics of patients who used the Teletriage Program during the pilot study, to explore safety and to report user acceptance and satisfaction. Methods: Cross-sectional study, including all adult patients affiliated to our health insurance attended to via this telemedicine service between January 18th and May 31st, 2019 (during pilot-study). Patients were followed-up for seven days to assess re-consultation to ED or unscheduled hospitalization. Results: 276 effective consultations occurred, corresponding to 241 patients, with a mean of age of 50 years, 68% (189) were women. Chief complaints were related to clinical issues (70%) and remaining (30%) were administrative problems. Only four patients were suggested a referral or face-to-face assessment. Rate of re-consultation to the ED was 18% (51) at seven days of follow-up, and the rate of unscheduled hospitalization was <1% (2), both with good clinical evolution. Patient satisfaction was 72.73%, and regarding acceptability, 66.12% stated that without this channel they would have attended to a face-to-face consultation and 64.02% that they would do so if their needs remained unmet. Conclusions: Implementing this new communication channel could be a useful and safe strategy to reduce unnecessary non-urgent consultations to the ED.

Keywords. eHealth, mHealth, teletriage, telehealth, telemedicine

1. Introduction

Non-urgent consultations to the Emergency Department (ED) contribute to a phenomenon known as overcrowding [1]. Evidence suggests that telemedicine plays an important role in evaluating patients that would otherwise attend to the ED. On the one hand, support by remote pre-hospital electrocardiogram for patients with suspected acute cardiovascular disease proved to avoid unnecessary hospitalizations and to confirm correct diagnoses, by-passing delays in the ED and therefore lowering mortality and costs [2–4]. Furthermore, the use of a telemedicine consultation program significantly decreased neuro-emergent stroke patient transfers from rural hospitals to

¹ Corresponding author: Santiago Andrés Frid. E-mail: santiago.frid@hospitalitaliano.org.ar

urban settings, while increasing stroke specialist reading of patient imaging studies within 3 hours of the onset of stroke symptoms [5]. This experiences, as well as previous evidence on the use of telecommunications as a strategy to reduce non-urgent consultations to the ED [6], have reported reducing costs through a safe intervention [7], with elevated acceptance by users [8].

Our institution implemented a virtual program for low complexity respiratory teleconsultations during the epidemiological outbreak of Influenza in 2018, which proved to be a safe and effective strategy [9]. In that project, we perceived unmet clinical and administrative needs, which prompted us to implement a broader program for mild symptoms and non-urgent clinical consultations for adult patients.

The objective of this article is to describe the characteristics of patients who used this service during the pilot study, as well as to explore patient acceptance and satisfaction with the use of this tool.

2. Methods

Observational, descriptive, cross-sectional study at Hospital Italiano de Buenos Aires, a high-complexity hospital located in Buenos Aires, Argentina, that has its health insurance called Plan de Salud (PS) with over 150,000 affiliates. Our average daily rate of consultations to the ED is 500, 50% of which are attended to at a low complexity walk-in area with an admission rate below 1%.

Teletriage program was implemented and made available during the pilot period, on business days between 4 P.M., and 8 P.M. Consultations were initiated by patients through their Personal Health Records (PHR) via desktop or mobile devices, and attended to and registered by physicians in the EHR, with a structured form designed ad hoc. Variables in this form were used as secondary databases for the retrospective data collection.

We included all adult patients affiliated to PS who performed a teleconsultation between January 18th and May 31st, 2019. We excluded all teleconsultations with failure to communicate (initiated but not answered), non-effective consultations (failure with video, audio or chat, or with interruptions to the communication that could not be by-passed through an effective phone-call) or those with no complete structured form in the HER. Patients were followed-up for seven days from the date of teleconsultation to assess re-consultations to the ED and hospitalization.

Additionally, we evaluated waiting times (measured as the difference between the moment that the patient checks in and the moment in which the physician initiates the communication), patients' acceptance of the tool (through two mandatory questions to users before teleconsultation) and patients' satisfaction (measured through Telehealth Usability Questionnaire -TUQ- adapted from English [10]).

Mandatory questions with multiple-choice options were: (1) If this communication channel did not exist, what would you do? (A- Attend to a face-to-face consultation: walk-in clinic, ED; B- I would not carry out any consultation; C- Solicit an appointment with a physician; D- Use another Telemedicine system) and (2) If you could not solve your chief complaint through this teleconsultation system, what would you do to solve it? (A- Attend to a walk-in-clinic or ED, B- I would not carry out any consultations, C- Solicit an appointment with a physician).

Quantitative variables are presented according to distribution as means and standard deviation (SD), or median and interquartile range (IQR) or 25-75 percentiles.

Categorical variables are presented as absolute frequency and relative frequency (percentage). Rates are presented as prevalence, with its 95% confidence interval.

This research project was approved by the institutional review board. Confidentiality was guaranteed. There were no potential risks for patients.

3. Results

There were 476 virtual consultations registered. After applying exclusion criteria, we had a total of 276 effective consultations with a complete structured form, pertaining to 241 patients (see **Figure 1**).

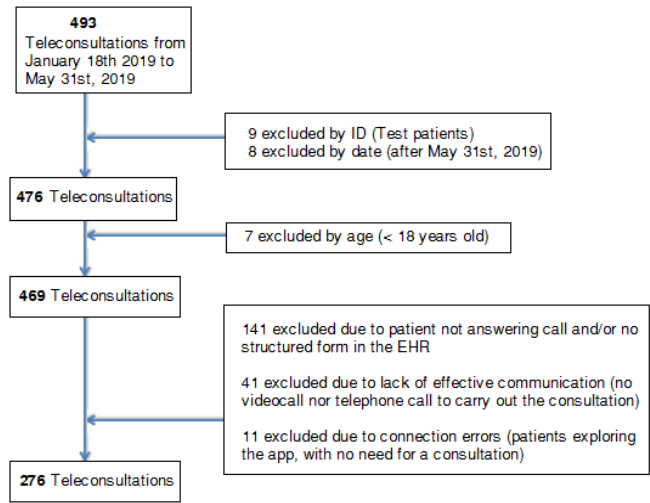


Figure 1– Flow chart of included patients

Service was offered for 90 days during the study period. Patients had a mean age of 50.46 years (SD 19.45), including patients from 18 to 97 years old, and 68.47% (189) were women. Data of care process was registered in EHR. Majority of chief complaints were clinical issues (detailed in **Table 1**) and solved through this channel. Administrative cases were related to drugs prescriptions or information, appointments, vaccines prescription, test results and medical information.

Table 1. Consultations and patient attention process

Chief complaint	n: 276
Administrative	29.34% (81)
Clinical	70.66% (194)
Respiratory	18.11% (50)
Infectious	17.75% (49)
Pain	13.76% (38)
Gastrointestinal	13.40% (37)
Gynaecological	3.98% (11)

Cardiological	2.89% (8)
Mental Health	0.77% (2)
Referral (ED, walk-in centre, doctor home visit)	1.44% (4)

The physician suggested a referral for a face-to-face assessment for only four patients (1.44%). At the seventh day of follow-up, the rate of re-consultation to ED was 18.47% (51), and the rate of unscheduled hospitalization was 0.72% (2), due to an asthmatic crisis and dehydration respectively. One patient had been referred to the ED while the other was already on her way to ED (during teleconsultation was carried out). Both had good clinical evolution.

In terms of waiting times, we found a median of 4 minutes (RIQ 9) and a mean of 7 minutes (SD 7.16).

As to patients' acceptance of this communication channel, 667 patients answered the mandatory questions (some never effectively checked in for a teleconsultation, and others experienced technical difficulties). 66.12% stated that without this channel, they would have attended to a face-to-face consultation and 64.02% that they would do so if their needs were not met with Telemedicine.

Regarding satisfaction assessment, 110 patients answered the TUQ (response rate of 23.45%). The percentage of patients that agreed 7/7 with the statements of the questionnaire were as follows: 'I am satisfied with the use of the system' (72.73%), 'I would use it again' (76.36%), 'Telemedicine is an acceptable way of receiving provision of healthcare' (73.64%), 'Teleconsultations are equivalent to face-to-face consultations' (28.18%), 'I am comfortable communicating with the physician via the Telemedicine system' (74.54%), 'I could see the physician as if I saw him in person' (52.73%), 'I could hear the physician when using the system' (57.27%) and 'I would recommend the Telemedicine system of HIBA to other people' (74.55%).

4. Discussion

This pilot study shows that implementation of an unscheduled teleconsultation program was a useful and safe strategy to meet patients' needs, representing an innovative and feasible experience for our institution, in compliance with recent suggestions by our Health Ministry for the development of Telemedicine in Argentina.

Additionally, based on emerging legal and moral consensus on hospitals' obligations towards its surrounding communities, Emergency Departments should not be treated as mere sources of admissions [11]. Therefore, management strategies such as this one are needed in order to improve the functioning of a collapsed system.

In consistence with our previous work, many chief complaints were administrative [9], which could be due to delays in patient attention in the outpatient setting. Our pilot also showed a high rate of satisfaction and acceptability. The lower percentages in patient satisfaction regarding how users saw and heard physicians were in line with the technical difficulties some patients experienced. This is mainly due to poor home Internet connections, which are highly prevalent in our country.

Lessons learned throughout this project will fuel improvements needed before implementing it at a larger scale. In consistency with evidence in the literature, we detected most usability issues in the initial phase of implementation of the telemedicine

system [12]. These issues prompted us to carry out a redesign of the tool and the process at the end of the pilot, with results that will be reported in the near future.

There are certain limitations to our study that need to be addressed. Real re-consultation rates can be underestimated since we did not take into account consultations to peripheral walk-in clinics that are part of our institution because we were unable to collect such information retrospectively. The analysis was also affected by effective teleconsultations not registered in the structured form in the EHR, due to the elevated number of physicians assigned to this project (higher possibility of errors in the process). The TUQ response rate was 23.45%, which could suggest a potential information or response bias, a denominator of self-selection or volunteer effect that has been reported in the literature [13], and it was adapted from English but not validated [14].

In accordance with the literature [15], we believe that part of our success was due to the existence of a local framework and the involvement of an interdisciplinary team, the activity of institutional communication channels, and the support provided by the Health Informatics Department, which allowed us to prioritize the caregiving process, maintaining efficacy of the workflow and improving providers' trust on the project.

References

- [1] D.H. Giunta, A.S. Pedretti, C.M. Elizondo, et al., Analysis of Crowding in an Adult Emergency Department of a tertiary university hospital, *Rev Med Chil* **145** (2017), 557–563.
- [2] N.D. Brunetti, N. Tarantino, G. Dellegrottaglie, et al., Impact of telemedicine support by remote pre-hospital electrocardiogram on emergency medical service management of subjects with suspected acute cardiovascular disease, *Int J Cardiol* **199** (2015), 215–220.
- [3] N.D. Brunetti, L. De Gennaro, M. Correale, et al., Pre-hospital electrocardiogram triage with telemedicine near halves time to treatment in STEMI: A meta-analysis and meta-regression analysis of non-randomized studies, *Int J Cardiol* **232** (2017), 5–11.
- [4] N.D. Brunetti, G. Dellegrottaglie, C. Lopriore, et al., Prehospital telemedicine electrocardiogram triage for a regional public emergency medical service: is it worth it? A preliminary cost analysis, *Clin Cardiol* **37** (2014), 140–145.
- [5] J. Whetten, D.N. van der Goes, H. Tran, et al., Cost-effectiveness of Access to Critical Cerebral Emergency Support Services (ACCESS): a neuro-emergent telemedicine consultation program, *J Med Econ* **21** (2018), 398–405.
- [6] J.R. Langabeer, T. Champagne-Langabeer, D. Alqusairi, et al., Cost-benefit analysis of telehealth in pre-hospital care, *Journal of Telemedicine and Telecare* **23** (2017), 747–751.
- [7] V. Tolia, E. Castillo, D. Guss, EDTITRATE (Emergency Department Telemedicine Initiative to Rapidly Accommodate in Times of Emergency), *J Telemed Telecare* **23** (2017), 484–488.
- [8] M.L. Barnett, K.N. Ray, J. Souza, et al., Trends in Telemedicine Use in a Large Commercially Insured Population, 2005–2017, *JAMA* **320** (2018), 2147.
- [9] S.A. Frid, M.F.G. Ratti, A. Pedretti, et al., Telemedicine for Upper Respiratory Tract Infections During 2018 Epidemiological Outbreak in South America, *Stud Health Technol Inform* **264** (2019), 586–590.
- [10] B. Parmanto, A.N. Lewis Jr, K.M. Graham, et al., Development of the Telehealth Usability Questionnaire (TUQ), *Int J Telerehabil* **8** (2016), 3–10.
- [11] Should Hospital Emergency Departments Be Used as Revenue Streams Despite Needs to Curb Overutilization?, *AMA Journal of Ethics* **21** (2019), E207–214.
- [12] S.O. Agnisarman, K.M. Chalil, K. Smith, et al., Lessons learned from the usability assessment of home-based telemedicine systems, *Appl Ergon* **58** (2017), 424–434.
- [13] C. Manterola, T. Otzen, Los Sesgos en Investigación Clínica, *International Journal of Morphology* **33** (2015), 1156–1164.
- [14] J.M. Ramada-Rodilla, C. Serra-Pujadas, G.L. Delclós-Clanchet, Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas, *Salud Pública de México* **55** (2013), 57–66.
- [15] Organización Panamericana de la Salud, *Marco de Implementación de un Servicio de Telemedicina*, DC : OPS, Washington, 2016.

The Renal Health Instagram: An Analysis of Comments

Geraldo Bezerra da SILVA JUNIOR, MD, PhD ^{a,b,1}, Marjan ASKARI ^c, Diovana Ximenes Cavalcante DOURADO ^a, Juliana Gomes Ramalho de OLIVEIRA ^a, and José Eurico de VASCONCELOS FILHO ^d

^a*School of Medicine, Public Health Graduate Program, University of Fortaleza, Fortaleza, Ceara, Brazil.*

^b*School of Medicine, Walter Cantidio University Hospital, Federal University of Ceara, Fortaleza, Ceara, Brazil.*

^c*School of Health Policy & Management, Department of Health Services Management & Organization, Erasmus University, Rotterdam, The Netherlands.*

^d*Application Nucleus of Information Technology, University of Fortaleza, Fortaleza, Ceara, Brazil.*

Abstract. Chronic kidney disease (CKD) is a Public Health problem affecting a considerable number of patients worldwide. CKD treatment is complex and requires patients' education. Based on this we have created technological tools, including an application for smartphones and a profile on Instagram (Renal Health) aiming to educate patients to self-monitor and cope with their disease, to increase adherence to treatment. In this study, we have analyzed the spontaneous feedbacks patients, and other people have posted on the Renal Health Instagram (comments) to investigate which information was needed the most by which types of patients. During the first 15 months since the release of this profile, there were 3380 followers, a total of 449 posts, with 36,079 "likes". Most of the followers were patients and parents, they gave spontaneous testimonials of their experiences with having kidney disease, and were thankful for the information provided and for the response we gave to their questions. From the analysis of the comments, we found that information on nutrition, physical activities and kidney transplant were the most discussed and valued. Our results also showed that the main use of this medium for dialysis and transplant patients was to share their experience. The Renal Health Instagram was considered a good digital platform of trusted information for both patients and the general population and also a space for sharing experiences in the context of kidney disease epidemics.

Keywords. Kidney disease, patient empowerment, internet, Instagram, education

1. Introduction

Since 1980 debates regarding health promotion are advancing in Brazil, with focus on the multi causality of diseases and the right to health, culminating in the creation of the Brazilian Unified Health System (SUS), which is now one of the largest Public Health Systems in the world, with free universal coverage [1]. Since the creation of the SUS,

¹ Corresponding Author: Prof. Geraldo Bezerra da Silva Junior, Programa de Pós-Graduação em Saúde Coletiva, Universidade de Fortaleza, Av. Washington Soares, 1321. CEP 60811-905. Fortaleza, Ceara, Brazil; E-mail: geraldobezerajr@unifor.br.

health education is valued, and patients are considered as crucial in the process of healthcare and the management of their care, to have an active role for achieving a good health condition [2].

This protagonism of patients in healthcare is crucial in the context of chronic disease, as in the case of chronic kidney disease (CKD). CKD is currently a worrisome Public Health problem, which affects approximately 11-13% of the general population. This reality is not different than in Brazil, where these rates have grown, having a 27% increase of patients on hemodialysis from 2012 to 2017, and in 2018 there were more than 130,000 people receiving dialysis [3]. The main causes of CKD are hypertension and diabetes, which are controllable, require patients' education and collaboration to change their habits. Therefore, educational measures are crucial, and technology can help.

Recent data shows that more than 70% of internet users claim to have searched for health information in the past year [4]. In Brazil, recent data from the Ministry of Planning, Development and Management showed that 74.9% of households have internet access, and 97% of the population has a smartphone [5]. Although there are many sources of information on the internet, many are neither trusted nor valid. Therefore, there is a need to create educational material, with a multidisciplinary team, which can be frequently updated [2, 6]. In this way, patients and people who connect with our social networks and other technological tools that we create can have easy access to qualified professionals, who cannot be so easily accessed in regions where access to health is limited, such as in poor and developing countries. In addition, one size does not fit all. Therefore, the aim of our study was two-fold: first, we aimed to analyze the effects of a recently created social network for patients with kidney disease (the Renal Health Instagram) through their spontaneous comments, using the internet and smartphones as a way of getting information for health education.

2. Methods

The Renal Health project started in 2015, aiming to develop technological tools for patients, and the public in general, to get information about kidney disease in Fortaleza, Brazil, including an application to help CKD patients in their treatment [6]. Although the fifth largest city in Brazil, Fortaleza, still faces several problems, including poverty, lack of sanitation, unemployment, and difficulties with access to health care, we have the possibility to offer technological tools to help people cope with chronic diseases, and they have easy access to the internet. All this makes health education essential to achieve better results in terms of decreasing the burden of chronic diseases. In the most recent official document about the mortality causes in Brazil, CKD appeared as the 10th most common cause of death in this region [5]. We created an Instagram profile and a YouTube channel to educate people, using not only text but also movies. We posted information about nutrition, medication use, physical exercises, and several other aspects of kidney diseases, prevention and treatment, on a weekly basis. The profile can be accessed through this link: https://www.instagram.com/renal_health/. We analyzed the spontaneous feedback people who follow the Renal Health Instagram profile have given from 1st June 2018 to 8th September 2019. Aside from analyzing all the feedback given in this period, including a quantification of the "likes", we specifically analyzed the posts that were most commented on, and what type of information was required the most by which type of kidney disease patients. An analysis of the testimonials and

acknowledgements was also done, in order to identify which groups of patients and people were most “reached” by the information provided by the Renal Health Profile so that in the future posts we could offer more tailored information based on these patients’ needs. We have analyzed all the posts of the profile (“Renal Health” Instagram), as well as each comment and feedback, and made a classification based on the subject. It consisted of a content analysis. Subsequently, we have made a characterization of the socio-demographical profile of the followers.

3. Results

Since the first publication in the Renal Health Instagram profile on June 1st, 2018, we have gathered a total of 3380 followers. In total there were 449 posts: 14 (3%) about medication, 63 (14%) about nutrition, 9 (2%) about excesses, 123 (17,39%) about kidney disease and comorbidities, 70 (15,59%) about treatment stimuli, 55 (12,2%) about kidney transplants and organ donations, and 35,78 % others. Virtually all posts received “likes” and comments. A total of 36,079 likes have been observed, and the post that has received the highest amount of “likes” (n=837) was one regarding organs’ donation.

Of a total of 1381 comments, these comments were performed by 469 profiles, of these are 80 (17%) men, 350 (74.6%) women and 39 (8.3%) profiles without gender, noting the prevalence of women among the comments made. One hundred eighty-five (185) comments were posted regarding congratulations and acknowledgements on the profile (Figure 1). People were thankful for the fact that our profile is interactive, and all questions are answered. We have identified 95 patients with kidney disease, mainly CKD. Of these, 73/95 (63.1%) spontaneously reported their illnesses in their comments, being on 60.2% dialysis, 21.9% transplanted patients, 6.8% glomerulonephritis, 5.4% kidney stones and 5.4% renal cysts. In total, 95 patients had posted spontaneous testimonials about their diseases and experiences with having CKD, and the difficulties they had faced. Also, 28 acquaintances of patients with kidney diseases had posted comments on our publications, most regarding questions to help them cope with the care of their children.

Table 1. Types of themes posted in the Renal Health Instagram, June – September 2019.

Posts category	Number of posts
Medication	14
Nutrition	63
Excesses	9
Kidney disease and comorbidities	123
Treatment stimuli	70
Kidney transplants and organ donations	55
Others	115
	449



Figure 1. Testimonials of patients following the Renal Health Instagram, showing the impact on their habits and lifestyle. On the right side, a testimonial showing a patient’s lifestyle change, and on the left, a tattoo with our project’s logo in a patient’s arm in a sign of gratitude and comprehension of the aims of the project.

The most commented posts were: 32 posts with comments on nutrition, followed by kidney transplants (with 29 comments), totalling 61 comments. From 90 comments that were posted in the groups of 45 patients on dialysis, there were 14 comments covering the questions on this topic, 68 comments of personal experiences on the subject, and finally eight comments on the importance of this information. In the group of 12 patients having kidney transplantation, 11 comments demonstrated their happiness after the transplant and their experiences, followed by one comment covering a question about their happiness after the transplant and their experiences. We have also identified other health professionals interacting with the profile, declaring that it was useful to help them to give advice to patients. And finally, this platform had crossed the borders, since we have also identified comments by people from other parts of the world, such as other South American countries, the Middle East and Europe.

Table 2. Types of comments from 45 dialysis patients posted in the Renal Health Instagram, June – September 2019.

Types of comments from dialysis patients	Number of posts
Questions on the topic	14
Personal experiences on the subject	68
About importance of this information	8

4. Discussion

The Renal Health Instagram showed to be a valuable digital platform for patients with kidney disease and surpassed the initial objective of the researchers, which was to provide information to patients and caregivers. Our results showed that methods of dialysis and treatment of chronic kidney disease were the most important information for the patients with dialysis and transplantation, followed by nutrition. Our results also showed that the main use of this media for dialysis and transplant patients was to share their experience.

Currently, Medicine and healthcare, in general, cannot be practised without technology and connecting with patients. According to Meskó (2017), there are very interesting new terms in Medicine, for example, the “e-patient”, which can be defined as the patient who is “empowered, engaged, equipped, enabled, equal, or expert”, and “Participatory Medicine”, which is “a model of cooperative health care that seeks to achieve active involvement by patients, professionals, caregivers, and others across the continuum of care... to increase patient satisfaction and improve the cost of care” [7].

The use of social networks, such as the Renal Health Instagram, by patients and the general population, as a way of getting trusted information for health education is benefic, and a real way of connecting patients and the healthcare team in a trusted relationship. Using these platforms, a broader audience can be reached, cross the borders, as the internet can be easily accessed from any part of the world. As technological tools may empower patients with information and abilities to cope with

chronic diseases, medical recommendations might be better followed and therapeutic goals achieved. It is expected that treatment's adherence increases, and therefore complications might decrease, and outcomes might improve. Clinical trials are needed to test this hypothesis.

There are few studies analyzing the impact of the use of Instagram on patients' health education. In a study regarding the use of this social media in the area of Orthopedics, posts on Instagram were predominantly from patients, and fewer were made from doctors [8]. These posts had in the majority of cases positive comments [8]. Another study evaluated the experience of patients using social networks regarding Twitter and Instagram to discuss their experience about a specific surgery (Gamma Knife stereotactic radiosurgery), and the authors evidenced that the main use of these media was to share patients' experience [9], which is accordance with our findings for transplant and dialysis patients. To the best of our knowledge, this is the first study to extract information need for specific types of CKD patients using comments of the provided general information for CKD patients in an internet social network. Future studies are needed to analyze bigger dataset to identify the specific information needs of CKD patients.

5. Acknowledgements

We are very grateful to the International Society of Nephrology (ISN) and the Edson Queiroz Foundation/University of Fortaleza for the financial support given to the Renal Health project. We also thank the Brazilian Research Council (CNPq) and the Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) for research grants provided for GBSJ and JGRO during the development of this study.

References

- [1] A. Massuda, T. Hone, F.A.G. Leles, M.C. de Castro, R. Atun, The Brazilian health system at crossroads: progress, crisis and resilience, *BMJ Glob Health* **3** (2018), e000829.
- [2] L.S. Fernandes, C. Calado, C.A.S. Araújo, Social networks and health practices: influence of diabetes online community on adherence to treatment, *Cien Saude Colet* **23** (2018), 3357-3368.
- [3] F.S. Thomé, R.C. Sesso, A.A. Lopes, J.R. Lugon, C.T. Martins, Brazilian chronic dialysis survey 2017, *J Bras Nefrol* **41** (2019), 208-214.
- [4] Brazil, Ministry of Planning, Development and Management, Pesquisa Nacional por Amostra de Domicílios (PNAD), *Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal*, IBGE, Rio de Janeiro, 2016.
- [5] Brazil, Ministry of Health, Saúde Brasil Estados 2018: uma análise de situação de saúde segundo o perfil de mortalidade dos estados brasileiros e do Distrito Federal, Ministério da Saúde, Brasília, 2018.
- [6] J.G.R. Oliveira, M. Askari, G.B. Silva Junior, R.A. Freitas Filho, J.E. Vasconcelos Filho, Renal Health: Na innovative application to increase adherence to treatment through self-monitoring for patients with CKD and provide information for the general population, *Kidney Int Rep* **4** (2019), 609-613.
- [7] B. Meskó, The guide to the future of Medicine: Technology and the human touch, Webcina Kft, Budapest, 2017.
- [8] H.S. Haeberle, N.I. Bartschat, S.M. Navarro, P.W. Rooney, J. Rosneck, R.W. Westermann, P.N. Ramkumar, Hip arthroscopy: a social media analysis of patient perception. *Orthop J Sports Med* **7** (2019).
- [9] R. Jacobs, A.V. Prabhu, E.A. Monaco, D. Tonetti, N. Agarwal, Patient perception of gamma knife stereotactic radiosurgery through twitter and Instagram. *Interdisciplinary Neurosurg* **13** (2018), 138-140.

Use of Social Media Apps by Nurses for Professional Purposes in Catalonia: Cross-Sectional Study

Antonio TORREJON^a, Octavi RODRIGUEZ-BLANCO^a and Miguel A. MAYER^{b,1}

^a *Nursing Association of Barcelona (COIB), Barcelona, Spain*

^b *Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute and Universitat Pompeu Fabra, Barcelona, Spain*

Abstract. Over the last few years, the use of social media mobile applications or apps (SMAs) has increased exponentially. The potential advantages of using these technologies by health professionals in clinical settings have been discussed many times. Considering that the nursing profession is the largest segment of the healthcare workforce in the majority of countries in the world, the impact of using these apps by these professionals is very relevant. The objectives of this study were, firstly, to determine if nurses were using SMAs professionally and the most frequent SMAs used and secondly, to find out if, among nurses, there is a need for training in the use of these mobile applications for professional purposes. The study is a descriptive cross-sectional study based on an Internet survey of 1,293 nurses in Catalonia (Spain). The average age of the respondents who had these apps installed on their mobile phones or tablets, was 43.12 (SD \pm 11.32) years old. WhatsApp was the most frequent SMAs used by nurses for professional purposes, and 79.2 % of nurses mentioned they used it several times a day. WhatsApp was the preferred SMAs for communicating with colleagues (31.2% of nurses) followed by Facebook (18.4%) and Twitter (11.3%). In contrast, the use of the SMAs was much less frequent as a means of communication with patients (7.2% in the case of WhatsApp). Nurses expressed their need for specific training in the use of these apps for professional purposes, indicating the interest and potential impact of the introduction of these technologies in clinical environments. The use of SMAs is quite common among nurses at the moment of the survey, and WhatsApp was the most popular one to support their professional activity. Based on the results of the survey, the Nursing Association of Barcelona (COIB) will consider the design of specific training activities in the use of SMAs in clinical settings.

Keywords. Social media, nurse role, survey, mobile health

1. Introduction

Over the last few years, the use of social media platforms and mobile health applications (apps) has increased exponentially. The potential advantages of using these tools by professionals in clinical environments have been discussed in several forums [1-3]. Currently, there are a lot of social media apps (SMAs), and Facebook,

¹ Corresponding Author: Miguel Angel Mayer, migueangel.mayer@upf.edu.

YouTube, Twitter, WhatsApp, LinkedIn, Pinterest, Tumblr or Instagram are among the most frequent ones[4-6], which are available to assist health professionals with many different tasks. Some of these tasks allow us to work in collaboration, to communicate with patients and other professionals, to perform patient management and monitoring, health education, training and research [7,8]. Nevertheless, it is not clear enough how to use these technologies in clinical settings considering the different aspects mostly related to ethical and legal aspects such as privacy of personal data and security [9].

Considering that the nursing profession is the largest segment of the healthcare workforce in the majority of countries in the world, including Spain [10, 11], the impact of the use of mobile applications by nurses, is highly relevant in the future of modern healthcare.

The objectives of this study were, firstly, to determine if nurses were using SMAs professionally and the most frequent SMAs used, and secondly, to find out if, among nurses, there is a need for training in the use of these mobile applications for professional purposes. The survey included questions about the use of SMAs and also health apps for professional purposes. In this paper, we present the results of the study related to the use of SMAs.

2. Methods

The study is a descriptive cross-sectional study based on an Internet survey of 1,293 nurses in Spain. LimeSurvey [12], an open-source platform, was used to gather the information of the survey online. In March 2018, an e-mail, including a link to the website with the survey, was automatically sent to the 26,907 registered nurses who had a contact e-mail in the Nursing Council of Barcelona (COIB) of Catalonia [13]. COIB is the second-largest Nursing Council in Spain with 34,327 registered nurses. The e-mail also included information about the purpose of the survey and that the information gathered would be analysed anonymously. After the first e-mail, two more e-mails were automatically sent as reminders within the following 30 days.

The survey consisted of 34 questions organised in three parts: (1) questions related to sociodemographic and professional information such as age, gender, workplace and professional activity or speciality, (2) questions about the use of social media and health apps for professional purposes and finally, (3) questions asking for the training needs for using SMAs for professional purposes. This paper is focused on describing the characteristics of the respondents and the use of SMAs for professional purposes. The description of the use of health apps for professional purposes was published in another journal [14]. Descriptive statistics were applied for all the items in the survey using the R programming language version 3.4.2 importing a .csv file with all the data. The governing board approved the study of the Nursing Association of Barcelona.

3. Results

The percentage of respondents to the survey who used SMAs for professional purposes installed on their smartphones or tablets was 46.4% (600/1,293). The remaining 693 respondents said that they did not use apps for professional purposes and were not included in the analysis. The average age of the respondents was 43.12 (SD \pm 11.32) years old, and there were no significant differences between the age of men and women

(t=-1.5078, $P = .1336$). The distribution of age was: 16% baby boomers (> 56 years old), 58% generation X (between 35-56 years old) and 26% millennials (< 35 years old). Concerning the professional area, most nurses worked in hospitals (48%) and primary care (24%). The remaining respondents worked in social healthcare (7%), management (5%), pre-hospital care (4%) and teaching/research (3%). Regarding the specialisation, the most frequent groups were general nursing (53%), medical-surgical (14%), paediatrics (8%) and family and community (7%).

Regarding the frequency of use of SMAs for professional purposes, WhatsApp was the most frequent SMAs used for professional purposes. 79.2 % of nurses reported using WhatsApp several times a day, followed by Facebook (28.7% several times a day), without significant differences comparing hospital and primary care use ($p=0.220$). Instagram, Twitter and Skype were not so frequently used for professional purposes. Overall, nurses preferred smartphones (91.2%) instead of tablets (8.8%) for managing SMAs professionally. Figure 1 shows the frequency of use of the seven most common SMAs as a means of communication, according to the respondents of the survey. WhatsApp was used preferably as a means of communication with colleagues (31.2% of nurses) as well as Facebook (18.4%) and Twitter (11.3%). Nevertheless, the use of these SMAs was much less frequent as a means of communication with patients (e.g. 7.2% in the case of WhatsApp). Most of them (81.8%) considered that it was necessary to receive specific training in the use of SMAs in professional environments, without significant differences among age groups and gender ($\chi^2=3.40$; $P=.347$).

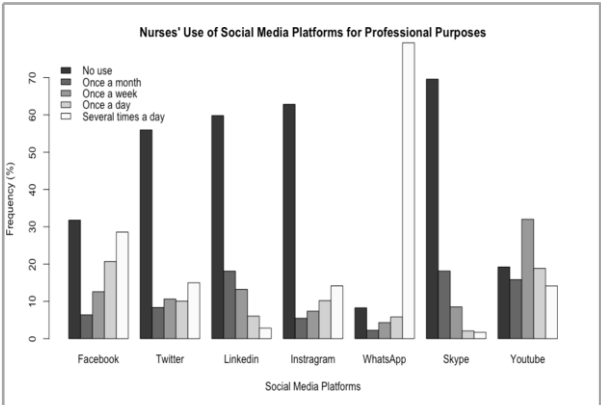


Figure 1. Use of social media apps by nurses for professional purposes

Most devices were less than two years old (77.0%), and 42.7% of the devices were less than one year old. As for the operating system, 51.1% had Android, and 45.7% had iOS. Table 1 shows the proportion of nurses that were using different SMAs as a means of communication with patients and professionals. WhatsApp and Facebook are the most frequently used SMAs as a means of communication with other professionals.

Table 1. The proportion of nurses that were using different social media apps as a means of communication with patients and professionals.

Social Media Apps	With patients (%)	With professionals (%)
Facebook	5.0	18.4
Twitter	2.2	11.3
LinkedIn	1.5	9.4
Instagram	1.2	4.5

WhatsApp	7.5	31.2
Skype	2.2	6.84
Youtube	1.8	3.0

4. Discussion

Over the last few years, the evolution of smartphones and other devices, as well as the applications developed, have evolved at an unprecedented rate. Our study shows that the use of SMAs is common among nurses (46.4%), being WhatsApp and Facebook the most frequently used apps to communicate with other professionals, as is the case of other health professionals such as medical doctors [6,15]. In contrast, the use of these apps is not frequent to communicate with patients. It could be of interest to know the content of these communications and the reasons why nurses use SMAs instead of other means of communication, considering that it is necessary to take into account different ethical and legal aspects when using these apps for professional purposes. Most likely, the ease of use and immediacy are the main reasons to use these SMAs.

The majority of nurses expressed their interest in receiving training in the use of mobile applications, indicating the potential impact of the introduction of these applications in clinical settings. It is essential to promote the training of nurses, and health professionals in general, who must be trained in the use of mobile applications, wearables or telemedicine applications. Currently, there is a need to analyse the specific advantages and drawbacks as well as consider all the aspects related to the use of these technologies among clinicians and it is necessary to analyse the ethical, legal, technical and security aspects to introduce SMAs in healthcare services [16-18]. In addition, the interoperability and the integration of SMAs in electronic health records should be evaluated very carefully to guarantee the quality, correctness and the compliance of all the legal and ethical requirements of each healthcare system. More experience is needed to determine the actual use and clinical usefulness of these technologies [19,20].

This study has some limitations, on the one hand, the number of participants in the survey was low compared with the number of nurses who potentially received the survey, and for this reason, the results could overestimate or underestimate the real use and interest of nurses in social media tools introducing a selection bias. On the other hand, the survey did not include all the types of SMAs available on the Internet that may be used by nurses. As far as we know, there are no other studies of the same characteristics in Spain or other countries that are focused only on nurses [1,6,15].

5. Conclusions

The use of SMAs is quite common among nurses at the moment of the survey. The most popular SMAs used by nurses to support their professional activity is WhatsApp, which was mostly used as a means of communication with other colleagues. Nevertheless, the use of SMAs is not as frequent as a means of communication with patients in clinical settings. Nurses expressed their need for specific training in the use and applications of SMAs for professional purposes, indicating the potential impact of the introduction of these technologies in clinical environments. Based on the results of the survey, the COIB will consider the design of specific training activities for nurses in the use of SMAs for professional purposes in clinical environments.

Acknowledgements

Special thanks to all the nurses of the Nursing Association of Barcelona (COIB) who participated in this study.

References

- [1] M.H. Mobasheri, D. King, M. Johnston, et al., The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study, *BMJ Innovations* **1** (2015), 174-181.
- [2] M.A. Mayer, A. Leis, A. Mayer, How medical doctors and students should use social media: a review of the main guidelines for proposing practical recommendations, *Studies in Health Technology and Informatics* **180** (2012), 853-857.
- [3] D.S. Eng, J.M. Lee., The promise and peril of mobile health applications for diabetes and endocrinology, *Pediatric Diabetes* **14**(4) (2013), 231-8.
- [4] J. Spencer, 65 + Social Networking sites you need to know, <https://makeawebsitehub.com/social-media-sites/>, MakeAWebsiteHub.com, 2019.
- [5] *Most popular mobile social networking apps in the United States as of September 2019, by monthly users* (2019), <https://www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/>, Statista, 2019
- [6] F.X. Marin, F. Garcia Cuyas, R. Reig-Bolano et al., Social Networking App Use Among Primary Health Care Professionals: Web-Based Cross-Sectional Survey, *JMIR Mhealth Uhealth* **6**(12) (2018), e11147.
- [7] E. Ozdalga, A. Ozdalga, N. Ahuja, The smartphone in Medicine: a review of current and potential use among physicians and students, *Journal of Medical Internet Research* **14**(5) (2012), e128.
- [8] J. Jackson, R. Fraser, P. Ash, Social Media and Nurses: Insights for Promoting Health for Individual and Professional Us, *The Online Journal of Issues in Nursing* **19**(3) (2014).
- [9] A.J. Barton., The regulation of mobile health applications, *BMC Medicine* **10** (2012), 46.
- [10] *Registered Health Professionals Statistics*, http://www.ine.es/en/prensa/epsc_2018_en.pdf, National Institute of Statistics, Spain, 2018.
- [11] N. Crisp, L. Chen, Global supply of health professionals, *New England Journal Medicine* **370** (2014), 950-957.
- [12] LimeSurvey, <https://www.limesurvey.org>
- [13] Col·legi Oficial d'Infermers i Infermeres de Barcelona (COIB), <http://www.coib.cat>
- [14] M.A. Mayer, O. Rodriguez-Blanco, A. Torrejón, Use of health apps by nurses for professional purposes: web-based study, *JMIR mHealth uHealth* **7**(11) (2019), e15195.
- [15] A. Nikolic, N. Wickramasinghe, D. Claydon-Platt, et al., The use of communication apps by medical staff in the Australian health care system: survey study on prevalence, *JMIR Medical Informatics* **9** (2018), e9.
- [16] C.L. Ventola, Mobile Devices and Apps for Health Care Professionals: Uses and Benefits, *Pharmacy & Therapeutics* **39**(5) (2014), 356-364.
- [17] T. Wyatt, P. Krauskopf, E-health and Nursing: using smartphones to enhance nursing practice, *Online Journal of Nursing Informatics* **16**(2) (2012).
- [18] D.M. Billings, There's an "app" for that: tips for preparing nurses for roles in mobile health, *Journal Continuing Education in Nursing* **46**(9) (2015), 390-391.
- [19] F.C. Day, M. Pourhomayoun, D. Keeves, A.F. Lees, Feasibility study of an EHR-integrated mobile shared decision-making application, *International journal of medical informatics* **124** (2019), 24-30.
- [20] M. Anshari, Redefining Electronic Health Records (EHR) and Electronic Medical Records (EMR) to Promote Patient Empowerment, *International Journal on Informatics for Development* **8**(1) (2019), 35-39.

Visual Representation of African Traditional Medicine Recipes Using Icons and a Formal Ontology, ontoMEDTRAD

Appoh KOUAME^{a1}, Konan Marcellin BROU^a, Moussa LO^b, and Jean Baptiste LAMY^c

^aLARIT-LARIMA, Institut National Polytechnique Félix Houphouët Boigny,
Yamoussoukro, Côte d'Ivoire

^bLANI, Université Gaston Berger, Sénégal

^cLIMICS, Université Paris 13, Bobigny, France, INSERM UMRS 1142, Sorbonne
Université

Abstract. In traditional West African medicine, most practitioners are illiterate for official languages. 1127 local languages exist in this region. The transmission of knowledge in this traditional medicine is notoriously oral. In order to enable practitioners to communicate and preserve this knowledge, we are building SysMEDTRAD, a traditional medicine management system, based on a formal ontology of traditional medicine called ontoMEDTRAD. To overcome the language and illiteracy barriers, we propose to combine realistic photographs of plants with formal iconic descriptions of plants and recipes. This paper presents the formal and automatized representation of traditional African medicine recipes.

Keywords. Ontology, Traditional medicine, ontoMEDTRAD, West Africa, Iconic composition, Non-verbal communication.

1. Introduction

Traditional medicine (TM), despite the progress of modern medicine (MM), still exists. Without seeking to replace or to copy the MM, and even less to be obliterated by it, the TM practiced in Africa and especially in its western part, persists and knows a renewed interest at the dawn of the 21st century. This resurgence is partly justified by the fact that the populations health care needs are insufficiently covered by the MM. In addition, health care costs in MM are often expensive and prohibitive for the disadvantaged social groups [1]. Without any discrimination, almost all social classes, for their health and medical needs, even if they use MM, continue to refer to TM. About 80% of the population commonly uses TM [15]. In West Africa, and particularly in rural areas, every inhabitant knows and uses the therapeutic properties of a number of plants. We are therefore talking about popular medicine. In this part of Africa, we count 15 countries and 1127 local languages [13]. Thus, TMP (Traditional medicine practitioners) face a huge language barrier. Moreover, the majority of them do not know how to read nor to write the official languages. They speak them with difficulty or not at all. Knowledge of TM is widely transmitted orally, when it is not kept secret. Academic and research institutions related TM are almost non-existent in these countries. Most of these TMP are illiterate (the illiteracy rate varies from 65% to 80% [8]) and belong to a wide linguistic and cultural diversity. Consequently, a visual

¹ Corresponding Author, e-mail: kgerappoh@gmail.com

approach can be more appropriated for supporting the collaboration between TMP and also the preservation of TM knowledge.

SysMEDTRAD [2] aims at addressing this problem. It includes ontoMEDTRAD, an ontology of West African TM associated with a visual approach. ontoMEDTRAD is divided in two modules: a terminological module ontoCONCEPT-Term, and an iconic ontoICONE. In a previous work [2], we proposed a formal method for representing medicinal plants with icons. Here, we extend this representation with realistic photographs and we propose iconic recipes of plant-based medicines, automatically generated from the ontology. We focused on malaria [8], one of the most morbid and deadly diseases in Africa.

2. Background

Concerning TM, Ayimdji A. [5] made some tools in Cameroon, but without a visual approach. An ontology for African TM [4] from G. Atemezang, has its object focused on the validation of knowledge managed by a system of agents. It does not address the fact that the TMP are mostly not literate. In Côte d'Ivoire, VetoMed [6] of K. Brou explores some visual aspects, but does not contain an ontology.

Regarding MM, LAMY J.B. *et al.* proposed an iconic language named VCM (Visualization of Concepts in Medicine) for representing symptoms, disorders, treatments and follow-up surveillances [12]. VCM is equipped with formal semantics defined in an ontology [11]. VCM aims at facilitating and speeding up the access to medical data and knowledge, rather than crossing the language barrier.

3. Materials and methods

We carried out field visits to TMP accompanied by translators, in Côte d'Ivoire and Senegal. We collected TM documents from NPTMP, the National Program of TM Promotion (Côte d'Ivoire) and some NGOs such as PROMETRA (Sénégal). Given the highly implicit and hidden nature of TM, we have conducted sensitization activities with 50 Ivorian TMP to increase their openness and their commitment to the project. The Abidjan-based NPTMP manager played a facilitating role in these activities alongside the TMP. We also visited 59 Senegalese TMP, mainly in the Saint-Louis region. From published works [1, 10, 3, 9, 8, 16], we extracted 31 traditional medicines for treating malaria with 28 recipes made of 22 distinct medicinal plants. We extracted realistic photographs of these plants from Wikipedia.

We followed an incremental and progressive general approach. First, we considered three use cases for our system: identifying the plants for a given recipe, preparing the medicine, and administrating it to the patient. We structured the pieces of information describing the medicines and the recipes in a formal ontology of TM. This ontology was manually modelled using the Protégé editor. We also collected various concepts from existing top ontologies (*e.g.* open biological ontologies (OBO), general formal ontology (GFO) and descriptive ontology in linguistics and cognitive engineering (DOLCE)) and domain ontologies (*e.g.* plant ontology (PO) [14]).

Second, for describing and identifying plants, we combined two approaches: realistic photographs showing the leaves and the fruits, and icons showing the main botanic criteria permitting the identification of the plant. Candidate criteria were

extracted from botanic sources [3], and then selected according to their discriminant capacity, as determined by the Weka Ranker and JRIP modules, as described in our previous works [2]. The resulting criteria were integrated into the ontology.

Third, plant icons are obtained by combining various iconems (*i.e.* small pictograms or colors), each corresponding to a specific value for a given criterion. Three strategies were used for visual representation, inspired by Peirce: analogy *i.e.* visual similarity, index *i.e.* semantic association and symbol *i.e.* arbitrary conventions [7]. Iconic recipes are obtained by the juxtaposition of one or several ingredients. Each ingredient is a part of a plant, and is represented graphically by the photographs and the icons of the plants, and a small icon representing the part of the plant used (*e.g.* leaves or roots). Iconic recipes are then enriched with icons representing the preparation and administration modes.

Finally, pictograms were drawn using the Inkscape vectorial drawing software. The icons and iconic recipes are generated as Scalable Vector Graphic (SVG) image files, automatically from the knowledge in the ontology, using Python scripts and ontology-oriented programming with the OwlReady module [11]. These scripts combine the pictograms and colorize them as needed excepted leaf for which green color is retained.

4. Results

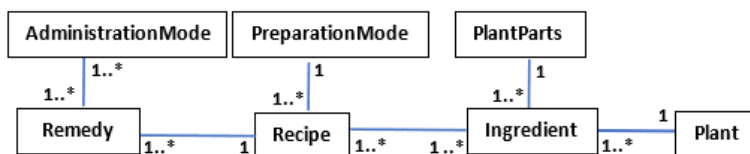


Figure 1: UML model of TM recipes and medicines.

The resulting ontology, previously described in [2], uses the *ALCHOIQ* Description Logic. Figure 1 shows the UML model of the plant-based medicine and recipe in ontoMEDTRAD. A medicine is characterized by a recipe and one or more administration modes (*e.g.* oral route, nasal route with fumigation, topic application during bath, ...). The recipe includes one or several plant-based ingredients and a preparation mode (*e.g.* decoction, infusion,...). Each ingredient is described by a plant and a part of a plant, using functional object properties. This ensures the uniqueness of ingredients.

The criteria retained for describing plant icons are given in Figure 2, with an example of icon for the plant *azadirachta indica* A. Juss.

Figure 3 gives an example of iconic recipe and its decomposition in icons. This recipe has two ingredients : leaves of *Azadirachta indica* A. Juss. and leaves of *Senna occidentalis* L. It is prepared as a decoction and the resulting medicine can be administered orally or in a bath. We included a total of 28 herbal recipes and 31 herbal antimalarial medicines in the current ontoMEDTRAD ontology.

5. Discussion and conclusion

In this paper, we proposed a formal method for the visual and iconic presentation of plant-based recipes in traditional West African medicine. This approach relies on a formal ontology of TM, namely ontoMEDTRAD.

Many ontologies exist for MM. However, it stands important differences in the way the patient is managed. In TM, diseases are detected by signs and symptoms that are clues and not necessarily by clinical examinations of MM. While remedies of TM are used in their almost natural state, MM often uses a synthetic product chain bound to a chemical industry. Unlike MM practioners, the TMP is at the same time a "physician", a "pharmacist" and a "nurse. Hospitalization and surgery are very rare in TM. Consequently, we did not reuse existing MM ontologies.

A few ontologies were proposed for TM [5, 4]. However, they are embryonic and they lack of medical validation. They also miss the information required for the visual presentation of plants as icons, e.g. they do not include the color of flowers and fruits. Therefore, we preferred to design our own ontology.

In addition to plant icons, we have included realistic photographs of the plants.

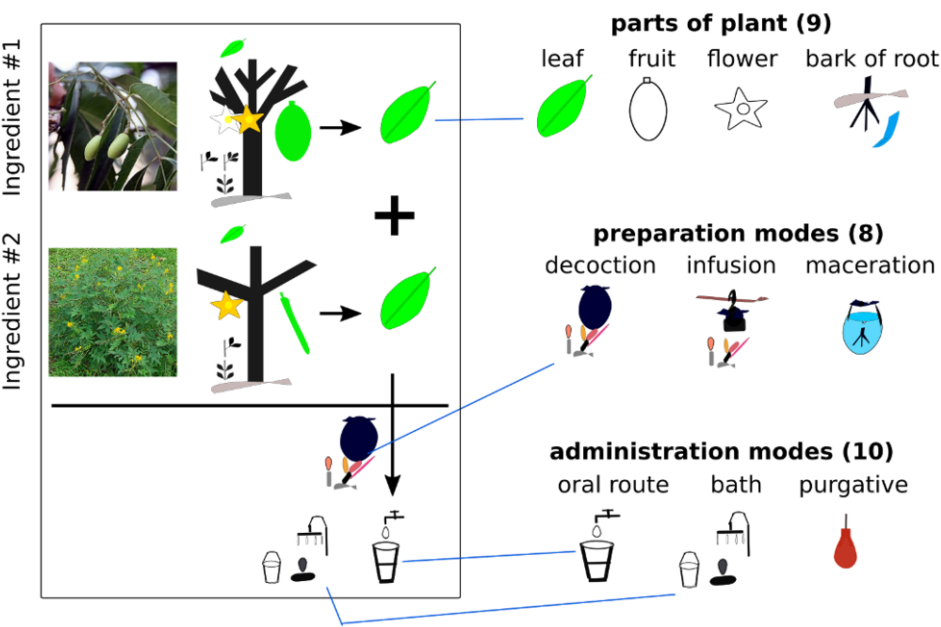


Figure 3: Example of iconic recipe (on the left) and examples of pictograms (on the right) (Numbers between parentheses indicate the number of distinct possible values for each criterion).

Both icons and photographs are visual presentation and description of the plant. They are highly complementary : the photographs allow an immediate recognition of the typical plant individuals, while the icons permit recognizing plant individuals that are significantly different from the photographs, *e.g.* due to the variability of the same plant species according to the various ecosystems, thanks to the botanic criteria that characterize the species. Pictograms were drawn by ourselves, however, they could be improved by a professional graphic designer without modifying the semantics and semiotics of our work. Both icons and photographs allow overcoming the linguistic barriers for TMP. A short training is probably needed before TMP fully understand our iconic language.

This work is part of a larger project aimed at the modernization and the valorization of West African TM. The main perspectives of the present work are : (1) the extension of the iconic and photographic presentation to other concepts such as symptoms, disorders, or mineral and animal medicinal resources; (2) the validation of the proposed visual presentation with TMP; (3) the integration of the visual presentation in a semantic wiki, allowing TMP to collaborate and to write and read medicinal recipes online.

References

- [1] Aké A ; Abrégé de Médecine et pharmacopée africaines ; Côte d'Ivoire ; SIPPI-NEI-CEDA; PP 1-150, vol 208, pp1-150 (2011)
- [2] K Appoh, B Lamy J , M Brou K and M Lo ; Modelization of recipe in african traditional medicine with visual ontology approach, iconic sketch ; Proceedings of e-Infrastructure and e-Services for Developing Countries, LNICST, scopus, Springer, Africomm 2016 ; PP 304-312 (2016)
- [3] Arbonnier M; Arbres arbustes et lianes de zones sèche de l'Afrique de l'Ouest ; 3 ème Edition QUAE 2009 ; ISBN MNHN 978-2-85653-626-1 ; PP 1-579 (2009)
- [4] G. Atemezing A and P Juan ; An Ontology for African Traditional Medicine ; Springer ; DOI 10.1007/978-3-540-85863-8_39. ISSN 1615-3871 ; pp 329-337 (2009)
- [5] Ayimdji A., S Koussoube, L P Fotso and K Balira O ; Using METHONTOLOGY to Build a Deep Ontology for African Traditional Medicine: First Steps; ARIMA; PP1-12 (2012)
- [6] M Brou K, C Bogui N, B Goore T and I Lokpo ; VetoMed , un système expert à base d'icônes pour la médecine vétérinaire traditionnelle ; Afrique Science, Vol.10, ISSN 1813-548X, pp 39-46 (2014)
- [7] N Carlos and Z T QINZ ; A Taxonomy of Representation Strategies in Iconic Communication, ScienceDirect Volume 70, Issue 8, pp 535-551 (2012)
- [8] Dasylyva B ; Contribution à l'étude de l'herboristerie traditionnelle sénégalaise: inventaire des plantes médicinales vendues dans les marchés de Dakar et contrôle de qualité sur 170 échantillons, UFR Pharmacie, UCAD, Sénégal., pp 1-100 (2001)
- [9] J Kerharo and G Adam J; La pharmacopée sénégalaise traditionnelle. plantes médicinales et toxiques ; Ed : Paris, Vigot frères, DL ; ISBN : 2-7114-0646-6 ; pp 1-200. (1974)
- [10] A. Konan ; Place de la médecine traditionnelle dans les soins de santé primaires, (Côte d'Ivoire), Université Toulouse III-Paul Sabatier, pp54-118 (2012)
- [11] J B Lamy and L F Soualmia; Formalization of the semantics of iconic languages : An ontology-based method and four semantic-power applications ; Knowledge-Based Systems 30(5):579-606 (2017)
- [12] J B Lamy, C Duclos; A Bar-Hen, P Ouvrard and A Venot, An iconic language for the graphical representation of medical concepts ; BMC Med Inform Decis Mak ; 8:16 (2008)
- [13] P Lewis, S Gary F and D Fennig C; Ethnologue : Languages of the world ; 19th edition
- [14] <https://www.ethnologue.com/> (2016).
- [15] M. Riichiro, K Kouji, S Toshinobu and K Yoshinobu; Construction and Deployment of a Plant Ontology ; EKAW 2000 : Knowledge Engineering and Knowledge Management Methods, Models, and Tools ; Computer Science., vol 1937 ; Springer, Berlin, Heidelberg ; pp 113-128, (2001)
- [16] WHO ; WHO traditional medicine strategy : 2014-2023, ISBN: 978 92 4 150609 0, pp 76 (2013)
- [17] P ZERBO, J MILLOGO R, O G NACOULMA and P VAN D ; Plantes médicinales et pratiques médicales au Burkina Faso : cas des Sanan ; Plantes médicinales ; Bois et forêts des tropiques, no307(1) ; pp 41-53, (2011).

What Do We Know About the Use of Chatbots for Public Health?

Elia GABARRON^{a,1}, Dillys LARBI^a, Kerstin DENECKE^b, Eirik ÅRSAND^a
^aNorwegian Centre for E-health Research, University Hospital of North Norway,
 Tromsø, Norway

^bBern University of Applied Sciences, Bern, Switzerland

Abstract. *Background and objective:* The number of publications on the use of chatbots for health is recently increasing, however to our knowledge, there are no publications summarizing what is known about using chatbots for public health yet. The objective of this work is to provide an overview of the existing scientific literature on the use of chatbots for public health, for which purpose have chatbots been used, and whether health-related outcomes have been reported. *Methods:* We carried out a literature review on this topic across 5 databases: Pubmed, EMBASE, PsychINFO, SCOPUS, and IEEE Xplore. Identified papers were classified according to their underlying technology, application area, and study design. *Results:* A total of 15 relevant papers were identified: eight of these papers were developmental studies that tested the feasibility or usability of a chatbot, and seven were interventional studies. All the interventional studies reported positive health-related outcomes associated with the chatbot use. *Discussion:* The first studies testing chatbots for public health seem very promising; however, there are various aspects that should be improved, including the chatbots' designs, studies' methods, and analysis and reporting of results. More high-quality studies and improved reporting of chatbots' use are needed.

Keywords. Chatbot, public health, mHealth, public health informatics

1. Introduction

Public health is defined as the branch of medicine concerned with the prevention and control of disease and disability, and the promotion of physical and mental health of the population on the international, national, state, or municipal level. One of the main concerns of public health is the growing burden of chronic diseases, which highlights the urgent need to create and make available interventions that can reach large populations at a low cost [1].

Fully automated self-help interventions based on chatbots could serve as highly cost-effective health promotion interventions for large groups of people. Chatbots are computer software programs based on statistical learning, statistical analysis and educational theories aiming at simulating a human conversation by text or voice message, that are easy to use and do not require familiarity with a specific user interface. Although the number of publications on the use of chatbots for public health

¹ Corresponding Author: Elia Gabarron. Norwegian Centre for E-health Research; P.O. Box 35 N-9038 Tromsø (Norway); E-mail: elia.gabarron@ehealthresearch.no.

is increasing, this is still a small and emerging field [2-4], and to our knowledge, there are no publications summarizing what is known about using chatbots for public health yet.

The objective of this work is to provide an overview of the existing literature on the use of chatbots for public health, what kind of underlying technologies have been used, for which purpose have chatbots been used, and which health-related outcomes have been reported.

2. Methods

In order to get an overview of how chatbots have been used for public health purposes, we carried out a literature review in July 2019. We searched for research articles involving the use or testing of a chatbot for public health purposes, reporting results that were published from 2015 until July 2019. The search strategy covered 5 databases (Pubmed, EMBASE, PsycINFO, SCOPUS, and IEEE Xplore) and comprised the following keywords: "chatbot" or "conversational agent" or "relational agent" or "dialogue system" or "Siri" or "Alexa" or "Cortana" or "Google Home" or "Google Assistant" in combination with "public health"; "health promotion"; or "health education". Relevant identified papers were then classified according to technology, application area, and study design.

3. Results

A total of 140 publications were identified in the initial literature review. After removing duplicates, and excluding publications for being out of scope, or not reporting any clinical outcome, we ended up with 15 publications.

Among the 15 included studies, 14 were based on an own developed chatbot, and only one study used a commercially available chatbot [5]. Four studies focused on voice-only chatbots [6-9], and 11 studies used a text-based or hybrid chatbot [5, 8-17]. Voice-only chatbots for public health purposes were tested only in feasibility or usability studies and were developed to provide counselling. Only four of the chatbots were connected or integrated with different social media or website, i.e. LINE, Wikipedia, WhatsApp, and WeChat [11, 12, 17, 18], and four more referred to chatbots connected to sensors or other Apps [10, 14, 15, 19].

Regarding the study design, eight of the papers showed results related to the development of a chatbot, including usability or feasibility studies [6-12, 14]. The chatbots from these eight developmental studies focused on motivating users to control or reduce their weight [11, 14], and to provide different kind of health information, such as HPV vaccination [6], medication [7], breastfeeding [10], diabetes [12], atrial fibrillation [8], and spinal cord injury [9]. Participants involved in these eight papers described the chatbots as easy to use [6, 8], engaging [11], and were positive towards using chatbots or were satisfied with them [7-10, 12, 14].

Seven of the included papers were interventional studies where health-related outcomes linked to chatbots' use were reported [5, 13, 15-19]. All of these interventional studies used a text-based or hybrid chatbot, and the bot was used to motivate and to provide counselling on healthy lifestyles and wellness [18, 19], to improve mental health [5, 16], to reduce reproductive health preconceptions [13], to

encourage loss of weight [15], and to motivate smoking cessation [17]. Table 1 summarizes the reported health-related outcomes in these interventional studies.

Table 1. Health-related outcomes reported by the 7 interventional studies

Reference	Health domain	Reported results (intervention duration)
Gardiner [19]	Healthy lifestyle/wellness	Chatbot group: significantly decreased alcohol consumption, and increased daily fruit consumption (1 month)
Souza [18]	Healthy lifestyle/wellness	100% of participants loved the experience of receiving incentives from the chatbot (1 week)
Ly [16]	Mental health	Chatbot group: participants who adhered to the intervention showed significant positive effects on psychological wellbeing and perceived stress (2 weeks)
Inkster [5]	Mental health	Significantly improvement in symptoms of major depression and a higher proportion of positive experiences among high chatbot users (3 months)
Jack [13]	Reproductive health	Chatbot use was significantly associated with a higher proportion of preconception risks being resolved (6 months)
L'Allemand [15]	Weight control/obesity	70% of participants had >4 interactions per day with the chatbot, and 37% of the daily challenges were completed successfully (4 months)
Wang [17]	Smoking cessation	Chatbot groups: much higher quit rates (2 months)

4. Discussion

Fifteen relevant papers on the use of chatbots for public health were identified in a literature review. Eight of them were developmental studies that only tested the feasibility or usability of a chatbot, while seven were interventional studies. All the interventional studies used a text-based or hybrid chatbot and reported benefits linked to its use, including significant increase in fruit consumption and wellbeing, improvement in depression symptoms, higher proportion of preconception risks solved, and higher smoking quit rates; as well as significant decrease in alcohol consumption, and positive experiences with interacting with the chatbot.

Nowadays, the role of chatbots in the public health system is still limited. Nonetheless, chatbots could provide auxiliary care through providing counselling and encouraging self-management of illness [3], adherence to treatment [20] or administrative services such as scheduling appointments [21], among others. With the continuous advancement in technology and behavioural sciences research, the competency and trustworthiness of chatbots are expected to be increased and promote its use in more complex and significant tasks in the health organizations [3]. First studies testing chatbots for public health seem very promising; however, there are various aspects that should be improved, including the chatbot design, study methods, and analysis and reporting of results.

Regarding the chatbot design, although there is evidence showing higher effectiveness of digital interventions when they are linked, integrated or delivered

through social media [22-24], only a few studies connected or integrated the chatbot to social media or other apps. The connection of chatbots to social media and integration of information from various apps and sensors would make chatbots more tailored to each individual, substantially enhancing the usefulness and motivation for the users to relate to chatbots – thus better possibilities to improve health parameters.

On the study methods, very few studies carried out an intervention, and many of these papers were of very short duration (down to 1 week), and consequently, increasing the likeliness that the positive effects were caused by the "Hawthorn effect" [25] – not the chatbot. More randomized trials should be done on this topic, and future research should increase the duration of the interventions to prove its sustained effect over time.

Better analysis and reporting of results are also needed. It is important to take into account that two of the papers that reported benefits, both of them testing the chatbot in the field of mental health, reported that the positive effect was only present in participants who adhered to the intervention, or that had higher use of it [5, 16]. Adherence to digital interventions is correlated with the characteristics of the innovation, and also with participants' interest towards it [26]. Therefore, and in order to move this field forward, it is relevant that researchers report, analyze and discuss usage metrics and determinants of attrition linked to the chatbots use in health interventions [26]. Security issues related to the chatbot are also a relevant issue that was unconsidered or unreported in the included papers, even though security is crucial in the health domain, and especially relevant for systems that are collecting or processing data from a user. Another security risk that needs further consideration is the possibility of chatbots to motivate the user to do actions that could be dangerous for the individual (e.g. obese person with type 1 diabetes who wants to lose weight, not eating due to encouragement from a chatbot – ending up in hypoglycemia – hospitalized or even dying). Who assumes responsibilities for these risks and the possible dangers, and therefore must be addressed by researchers?

Our literature review has covered five databases; however, we might have missed additional relevant publications on the use of chatbots for public health. Future reviews should expand the scope of the search and incorporate research projects that are being carried out currently, in order to have a better overview of the potential different uses and benefits that chatbots could have for public health in the near future.

Although there are promising results, there is not enough evidence supporting the use of chatbots for public health yet. More high-quality studies and improved reporting of chatbots' use are needed.

References

- [1] *Global action plan on physical activity 2018-2030. More active people for a healthier world*, <https://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf>, World Health Organization, 2018.
- [2] K. Denecke, E. Gabarron, R. Grainger, et al., Artificial intelligence for participatory health: applications, impact and future implications, *Yearb Med Inform* **28**(1) (2019), 165-173.
- [3] L. Laranjo, A.G. Dunn, H.L. Tong, et al., Conversational agents in healthcare: a systematic review, *J Am Med Inform Assoc*, **25**(9) (2018), 1248-1258.
- [4] K. Denecke, M. Tschanz, T. Dorner, et al., Intelligent Conversational Agents in Healthcare: Hype or Hope?, *Studies in Health Technology and Informatics* **259** (2019), 77-84.

- [5] B. Inkster, S. Sarda, and V. Subramanian, An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental wellbeing: Real-world data evaluation mixed-methods study, *JMIR mHealth and uHealth* **6(11)** (2018), e12106.
- [6] M. Amith, A. Zhu, R. Cunningham, et al., Early Usability Assessment of a Conversational Agent for HPV Vaccination, *Studies in health technology and informatics* **257** (2019), 17-23.
- [7] R.F.L. Azevedo, D. Morrow, J. Graumlich, et al., Using conversational agents to explain medication instructions to older adults, *AMIA Annu Symp Proc* (2018), 185-194.
- [8] E. Kimani, T. Bickmore, H. Trinh, et al., A smartphone-based virtual agent for atrial fibrillation education and counseling, *16th International Conference on Intelligent Virtual Agents*, Springer Verlag, 2016.
- [9] A. Shamekhi, H. Trinh, T.W. Bickmore, et al., A virtual self-care coach for individuals with spinal cord injury, *18th International ACM SIGACCESS Conference on Computers and Accessibility*, Association for Computing Machinery Inc, 2016.
- [10] H.L.N. Himanshi and I.M. Karunathilake, A model for the development of a chatter bot/chat bot for field of public health using the example of automated online breast feeding helper, *International Journal of Advanced Trends in Computer Science and Engineering* **8(1)** (2019), 300-302.
- [11] C. Huang, M. Yang, C. Huang, et al., A Chatbot-supported Smart Wireless Interactive Healthcare System for Weight Control and Health Promotion, *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2018.
- [12] S. Hussain and G. Athula., Extending a Conventional Chatbot Knowledge Base to External Knowledge Source and Introducing User Based Sessions for Diabetes Education, *32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, 2018.
- [13] B. Jack, T. Bickmore, M. Hempstead, et al., Reducing Preconception Risks Among African American Women with Conversational Agent Technology, *J Am Board Fam Med* **28(4)** (2015), 441-51.
- [14] T. Kowatsch, D. Volland, I. Shih, et al., Design and evaluation of a mobile chat app for the open source behavioral health intervention platform mobile coach, *12th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2017*, Springer Verlag, 2017.
- [15] D. L'Allemand, C.H. Shih, K. Heldt, et al., Design and interim evaluation of a smartphone app for overweight adolescents using a behavioural health intervention platform, In: 9th Fribourg Obesity Research Conference (FORC)-2017, *Obesity Reviews* **19** (2018), 102.
- [16] K.H. Ly, A.M. Ly, and G. Andersson, A fully automated conversational agent for promoting mental wellbeing: A pilot RCT using mixed-method, *Internet Interv* **10** (2017), 39-46.
- [17] H. Wang, Q. Zhang, M. Ip, et al., Social Media-based Conversational Agents for Health Management and Interventions, *Computer* **51(8)** (2018), 26-33.
- [18] M. Souza, T. Miyagawa, P. Melo, et al., Wellness programs: Wearable technologies supporting healthy habits and corporate costs reduction, *19th International Conference on Human-Computer Interaction, HCI International 2017*, Springer Verlag, 2017.
- [19] P.M. Gardiner, K.D. McCue, L.M. Negash, et al., Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial, *Patient Education and Counseling* **100(9)** (2017), 1720-1729.
- [20] B., Chaix, J.E. Bibault, A. Pienkowski, et al., When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot, *JMIR Cancer* **5(1)** (2019), e12856.
- [21] A. Palanica, P. Flaschner, A. Thommandram, et al., Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey, *J Med Internet Res* **21(4)** (2019), e12887.
- [22] J.M. Petersen, I. Prichard, and E. Kemps, A Comparison of Physical Activity Mobile Apps With and Without Existing Web-Based Social Networking Platforms: Systematic Review, *J Med Internet Res* **21(8)** (2019), e12687.
- [23] E. Gabarron, E. Arsand, and R. Wynn, Social Media Use in Interventions for Diabetes: Rapid Evidence-Based Review, *J Med Internet Res* **20(8)** (2018), e10303.
- [24] E. Gabarron and R. Wynn, Use of social media for sexual health promotion: a scoping review, *Glob Health Action* **9** (2016), 32193.
- [25] J. McCambridge, J. Witton, and D.R. Elbourne, Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects, *J Clin Epidemiol* **67(3)** (2014), 267-77.
- [26] G. Eysenbach, The law of attrition, *J Med Internet Res* **7(1)** (2005), e11.

Section 4

Precision Medicine and Public Health

This page intentionally left blank

A Stroll Along the Erroneous ePrescription Interactions Within the Belgian Pharmacy

Sven VAN LAERE^{a,1}, Ronald BUYL^a and Marc NYSSSEN^a

^aDepartment of Public Health (GEWE), Vrije Universiteit Brussel, Belgium

Abstract. After the national introduction of the ePrescription in Belgium to the broader public in 2014, community pharmacists are still not satisfied completely. Reasons are (1) low acceptance of the implementation due to slow systems, (2) high reported downtimes, and (3) alert fatigue due to technical and incomprehensible error messages. Therefore, we investigated which technical errors occur in the flow of handling an ePrescription and how these errors can be avoided. A cross-sectional design was used to capture interactions of all national community pharmacists connected to the national eHealth platform on a randomly chosen working day. Per interaction, the number of errors made was observed. In total, 567,883 interactions were registered and analyzed, of which the getPrescription interaction, to download the ePrescription from the national server, was most prevalent ($n = 196,433$; 37.21%). A difference of 14,961 interaction calls was observed without reaching a final state (delivered or undelivered). Reasons for these differences are repetitive calls for obtaining the ePrescription or by trying to obtain the prescription again when this should no longer be possible (e.g., when an ePrescription is already delivered or archived). When looking at the markAsDelivered, markAsUndelivered or markAsArchived interactions, most of the technical handling errors are due to attempts that generate not allowed state transitions. Most of these incorrect state transitions could have been avoided by maintaining the state diagram in the pharmacist's system to enforce legal transitions and by training the community pharmacists to handle ePrescriptions appropriately.

Keywords. Electronic prescribing, community pharmacy services, eHealth

1. Introduction

The Belgian electronic prescription (ePrescription) system consists of the typical 3Ps model between prescriber, patient and pharmacist [1]. The electronic prescriptions are securely transferred from the prescriber to the pharmacist, selected by the patient. Typically, from the back-office of the pharmacist, this flow is extended to a price-fixing service that is responsible for the billing to the patients' health care insurances in the name of the pharmacies, since a third-party payment system applies in the Belgian healthcare system [2].

In Belgium, the ePrescription works in a decentralized way where the ePrescription is stored on a central database server, in an encrypted manner. The prescribers send their prescriptions in XML-coded format to this national server, whereafter the patient may visit his pharmacist of choice to get his prescription delivered. Currently, the patient still

¹ Sven Van Laere, Department of Public Health, Vrije Universiteit Brussel, Laarbeeklaan 103, 1090 Brussels, BELGIUM. Tel: +3224774424; Fax: +324774000; Email: Sven.Van.Laere@vub.be

receives a paper proof of the electronic version, which possibly will vanish after the dematerialization phase where the government wants to avoid unnecessary paper use. Currently, the paper-based proof of electronic prescription serves as an access token and feedback mechanism for both the prescriber and the patient so that they can verify what was prescribed. On top of this paper proof, a unique Recip-e ID (RID) barcode is printed, identifying the electronic prescription in the Recip-e database so that the pharmacist can easily retrieve the ePrescription from the national database, by scanning the barcode.

However, even after five years of successful deployment, enabling ambulatory care providers prescribing and delivering of ePrescriptions, care providers are still unsatisfied with this system in place. Closed fora for pharmacists in Belgium[4] report about (1) low acceptance of the implementation due to slow systems, (2) high reported downtimes of the eHealth services, and (3) a raised alert fatigue due to error messages that are written in a technical language that are uneasy to understand for the pharmacist.

The Belgian ePrescription follows the state diagram represented in Figure 1. It corresponds to the “normal workflow” related to handling medical prescriptions in a community pharmacy. Initially, when the prescriber submits an ePrescription, this prescription is in the state *undelivered*. When the community pharmacist retrieves the ePrescription of the national server via the *getPrescription* interaction, the state changes into *inProcess*. When the pharmacist does not have the product in stock, the pharmacist should trigger a *markAsUndelivered* interaction so that the patient can get his drug delivery at another pharmacist. The status is then returned to *undelivered*. When the state of the ePrescription is *inProcess*, the community pharmacist might decide to revoke the prescription, when he, for example, thinks the prescribed medication can be harmful or irrelevant for the patient, by the *revokePrescription* interaction that sets the state of the prescription at *revoked*. The ePrescription will now no longer be available for other pharmacists. When the pharmacist decides to fill the prescription, he can trigger a *markAsDelivered* interaction, putting the state at *delivered*. Eventually, the ePrescription will be sent from the pharmacy that made the delivery to a price-fixing service, thereby informing Recip-e using the *markAsArchived* or *markAsArchivedBis* interaction, putting the prescription in the end state *archived*. All of these interactions use a RID argument to indicate on what ePrescription they want to perform their action.

The low perceived acceptance, high perceived downtimes and growing alert fatigue raised the question of investigating the reason for this. Therefore, we investigated which technical errors in the flow of handling ePrescriptions happen by searching errors in the audit trail of the interactions performed on ePrescriptions and how these errors can be avoided.

2. Methods

In this study, we used a cross-sectional study design where all ePrescriptions’ interactions were taken under the loop to observe the audit trail of interactions taking place at the pharmacy. We, therefore, chose to observe a random day in the delivery process, i.e. Friday, September 6th, 2019. Experiences of the Recip-e organization show that we have to multiply this number by about 25 times to obtain an estimate of the monthly reported number of interactions that take place. The reason for not multiplying

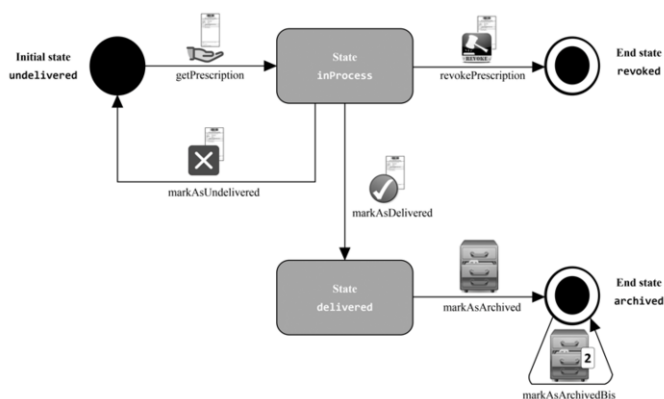


Figure 1. State diagram of Belgian ePrescription when received by the community pharmacist

this amount by 30 is that during weekends, Belgian citizens require fewer medications to be prescribed or filled than during weekdays.

The types of errors that occurred were identified and quantified by reporting the number and percentage of occurrences within the collected sample. The technical errors occurring were categorized as errors to be avoided by a good software implementation or as errors to be avoided by the community pharmacist user (by memorizing the state diagram presented in Figure 1 and actions taken).

3. Results

In the collection of the sample of September 6th 2019, a total number of 527,880 digital interactions of the community pharmacist were observed (see Table 1). Of this total number of interactions, 196,433 (37.21%) were *getPrescription* interactions, where the pharmacist tried to retrieve an ePrescription. When observing the number of interactions bringing the ePrescription from the initial undelivered state to the revoked state or the delivered state, we observed respectively 11,632 *markAsUndelivered* (2.21%), 39 *revokePrescription* ($< 0.01\%$) and 169,801 *markAsDelivered* interactions (32.17%). The sum of these interactions adds up to 181,472 interactions or 92.38% of the total number of *getPrescription* interactions. So, we found a difference of 14,961 ePrescription calls that requested the download of an ePrescription without being handled in a final form. A possible explanation lies in a total of 2,533 repetitive *getPrescription* calls where the ePrescription was already in a state no longer allowing to be fetched again (states archived, delivered, inProcess). The other differences ($n = 12,408$) are to be explained by the further manual processing of the prescription like they used to do when a paper-based prescription was used.

3.1. Incorrect *getPrescription* interactions

When observing the reason for these 2,533 repetitive *getPrescription* interactions, we observed that the highest number of repetitive calls occur when the ePrescription is already archived ($n = 1,547$; 60.60%). This type of error can be explained by the fact that a community pharmacist archives his prescriptions digitally towards a price-fixing

Table 1. Interactions performed on one day, and the number of errors made per type

Interaction	Total nr of interactions	Total nr of pharmacies	Nr of errors	Percentage of errors (%)
getPrescription	196,433	4,688	2,553	1.30
markAsDelivered	169,801	4,667	1,419	0.84
markAsUndelivered	11,632	3,645	534	4.59
markAsArchived	147,901	4,211	1,302	0.88
markAsArchivedBis	2,074	890	-	-
revokePrescription	39	26	-	-
TOTAL	567,883	4,859	5,808	1.02

service and tries to retrieve the digital content a second time. However, since the content has already been archived, its retrieval is no longer possible. Also, copies of the paper proof of prescription that were already delivered and archived at a different pharmacist will result in this incorrect state transition. This type of transition was observed in a total of 367 community pharmacists, of which 279 community pharmacists took care of the archiving of ePrescriptions themselves. It is expected that the user could avoid errors emanating from the user’s pharmacy as the user’s system should keep track of the handling of electronic prescriptions processed at this pharmacy.

3.2. *Incorrect markAsDelivered, markAsUndelivered or markAsArchived interactions*

The remainder of the incorrect calls deal with the *markAsDelivered* (n = 1,419), *markAsUndelivered* (n = 534) and *markAsArchived* (n = 1,302) interactions. Comparing this number of incorrect calls to the total number of interactions made of that type, we respectively obtained 0.84%, 4.59% and 0.88% of incorrect calls (see Table 1).

The main reason for all of these incorrect interactions is again the triggering of an incorrect action coupled to a not allowed state transition: (1) 1,005 of the 1,419 incorrect *markAsDelivered* interactions (70.82%); (2) 518 of the 534 incorrect *markAsUndelivered* interactions (97.00%); and (3) 1,302 of the 1,302 incorrect *markAsArchived* interactions (100.00%). All these incorrect state transition attempts were rejected but could have been avoided by the user when the state diagram, presented in Figure 1, would be memorized by the pharmacist and would be correctly enforced by the pharmacy system.

Other incorrect calls that were made could have been avoided by generating a correct call, i.e., at least giving a RID code as argument when a *markAsDelivered* call was generated (414 out of 1,419 calls; 29.18%) or using a correct RID code when a *markAsUndelivered* interaction was called (16 out of 534 calls; 3.00%). A good software implementation could additionally help verify these checks.

4. Discussion and conclusion

In this research, we showed that on an average working Friday, the Belgian community pharmacists jointly make 527,880 interaction calls with the national database in order to retrieve, (un)deliver, or revoke the ePrescription. This total amount does not include other calls being made by the prescriber to generate and store the ePrescription in a secure way to the server via the national eHealth platform. Five years after national deployment was initiated, still lots of problems related to slow responses, high downtimes and alert fatigue due to technical and incomprehensible error messages are perceived by the community pharmacist. This cross-sectional research tried to find reasons, by looking at the digital

interactions and errors made during interaction that takes place at a community pharmacy with the ePrescriptions on a randomly chosen working day.

About 1.10% of the interactions of pharmacists resulted in attempting to make a technically invalid interaction with the national database. Community pharmacists then typically get a quite technical message of their software stating something about the current status of an ePrescription. However, most of the community pharmacists do not understand these error messages, that in the end trigger their alert fatigue. An example of such a message is “Retrieval of ePrescription failed - [ERR100028] ePrescription BEP1PDT1F5SP is archived”. Here an incorrect state transition was requested.

Most of these error messages are easy to avoid, keeping the allowed state transitions in mind. This status follow-up should be tracked as well by the pharmacy software, but often this is not the case. However, when a patient presents the Belgian proof of ePrescription to a pharmacist, who does not know of earlier interactions in a different community pharmacy, these errors may seem odd and not appropriate. All too often, the pharmacist still tries to execute this prescription, possibly in its paper-based form [2].

Clearer alerts could help in the prevention of this type of errors. Also, clear instructions for the community pharmacist of what state transitions are allowed could help to this. In Belgium, future pharmacists are trained to work with digitalized systems by doing internships in different pharmacies. However, when this state diagram is not present in the mindset of tutoring community pharmacists, the tutor during the internship will not be able to guide the intern correctly.

An additional *verifyStatus* interaction might help to limit certain incorrect interactions of the community pharmacist by disabling them in the interface. However, a multiplication of the number of interactions is not desirable, since nowadays we are confronted with a high amount of downtimes due to technical issues at the level of our national eHealth platform. The finding of these high unavailability rates, due to downtime of the national services, was first reported by Suykerbuyk et al. [5] when reporting about the perceptions of the patient about the national ePrescription.

References

- [1] P. Esmaeil Zadeh and M.C. Tremblay, A review of the literature and proposed classification on e-prescribing: Functions, assimilation stages, benefits, concerns, and risks, *Res Social Adm Pharm* **12** (2016), 1-19. doi: 10.1016/j.sapharm.2015.03.001.
- [2] S. Van Laere, P. Cornu, E. Dreesen, J. Lenie, and R. Buyl, Why do Belgian Community Pharmacists Still Treat Electronic Prescriptions as Paper-Based?, *J Med Syst* **43**(11) (2019), 327. doi: 10.1007/s10916-019-1456-5.
- [3] M. Nyssen and M. Buckens, Deployment of Electronic Prescriptions in Belgium, *Proceedings of the Seventh International Conference on eHealth, Telemedicine, and Social Medicine, eTELEMED*; Lisbon Portugal, 2015.
- [4] D. De Proot, *Meer dan 4.000 apothekers in besloten Facebookgroep. de Apotheker*, <http://www.deapotheker.be/actualiteit/meer-dan-4-000-apothekers-in-besloten-facebookgroep/article-normal-29849.html>, 2017.
- [5] L. Suykerbuyk, M. Robbrecht, S. De Belder, H. Bastiaens, W. Martinet, and H. De Loof, Patient Perceptions of Electronic Prescriptions in Belgium: An Exploratory Policy Analysis, *Pharmacy, Communication* **6**(4) (2018), 130. doi: 10.3390/pharmacy6040130.

A Structured Measurement of Highly Synchronous Real-Time Ballistocardiography Signal Data of Heart Failure Patients

Nico JÄHNE-RADEN^{a,1}, Udo BAVENDIEK^b, Henrike GÜTSCHLEG^a, Ulf KULAU^c,
Stephan SIGG^d, Marie WOLF^a, Tanja ZEPPERINICK^b and Michael MARSCHOLLEK^a

^a *Peter L. Reichertz Institute for Medical Informatics*

of TU Braunschweig and Hannover Medical School, Germany

^b *Department of Cardiology and Angiology – Hannover Medical School, Germany*

^c *DSI Aerospace, Bremen, Germany*

^d *Department of Communications and Networking, Aalto University, Finland*

Abstract. Ballistocardiography (BCG) has gained more attention due to the fundamental goal of medical intervention in diagnostics and follow-up. BCG is particularly suitable for the study of heart failure, which a recent study has shown. The results of this working group shall be validated and reproduced with another study trial. Therefore, acceleration sensor prototypes will be placed on various parts of the patient's body and be connected to a computer unit, which allows a high data quality and high signal resolution. A temporal shift of only 20 ns ensures real-time measurement of BCG parameters. The reference measurement will be done with a 12-channel ECG. The study will include patients with heart failure. All conducted tests take place as part of the diagnostic-therapeutic routine. The only change in the procedure concerns the additional equipment with the measuring sensors. The results will be the validation of the data from the other working group, as well as the information about the choice of sensors and clock frequency, the measuring points and the needed features for early detection of heart failure in BCG signals.

Keywords. Ballistocardiography, Seismocardiography, Heart Failure

1. Introduction

In recent years, the ballistocardiography has gained more attention [1–10], with the fundamental goal of medical intervention, for example, the diagnosis or follow-up.

However, ballistocardiography or seismocardiography [4,5] is particularly suitable for the study of heart failure (HF), as new studies confirm [10].

With the prospective study described here, the previous explorative results of the scientific society are to be reproduced, and this validates their validity.

¹ Corresponding Author, Nico Jähne-Raden, E-mail: Nico.Jaehne-Raden@plri.de.

The determination of the course of development of heart disease or long-term measurements of cardiac parameters are currently still difficult to implement and can only be determined with the help of medical visits or measurements of only 24 hours. However, there are diseases that have a much slower course or can appear at irregular intervals and are therefore unnoticed or measured within 24 hours.

Ballistocardiography should be a simple and inexpensive solution, but at the same time efficient and highly specialized [4,5,9,11,12]. Thus, it is possible to determine the blood pressure with the aid of the measured pulse transit time (time difference of the same signal between various measurement points on the subject's body), without the patient being awakened by inflating a cuff around the arm at night or interfering with his activities during the day, thereby distorting the result [13]. Since such a system is very small and has a battery with several days of running time, measurements of more than 24 hours (as a single measurement) can be made to reduce the disease deterioration of e.g. Cardiac insufficiency or high blood pressure (even over a longer period, for example, over months) [4,5].

At present, there are only a few ballistocardiographic measurements performed in everyday life or with cardiac patients. Data on the signal morphology of cardiac insufficiency have only been collected in an exploratory fashion [1]. With this prospective study planned, we then want to validate these exploratory results. Furthermore, in literature found setups raise further questions, so the technical measurement setup from the point of view of the working group is not optimal (choice of sensor, choice of clock frequency, choice of measurement position).

The overall objective of this prospective study is the validation of the results from the published results [10] for the measurement of ballistocardiographic or rather seismocardiographic signals in cardiac patients and their usability for heart failure diagnostics and/or course analysis, through the use of a highly specialized sensor measurement system consisting of body-worn acceleration sensor.

Secondary goal is to find out what features are needed for early detection of heart disease (e.g., Graph Similarity Score or other correlation) and what the signals of such diseases look like in a ballistocardiographic record.

2. Methods

The BCG sensor system is intended for the measurement. Components include the core unit, the FPGA processor board, and seven suitable triaxial acceleration sensors (Accelerometer). For the reference measurement, a 12-channel ECG is read out via the analog output parallel to the acceleration signals. This is a common method to be able to compare the two signals to each other and to provide information when in the BCG signal the characteristic features should appear.

The measuring system allows a high data quality (low signal-to-noise behavior) and a high signal resolution (2g scaling at 16 bit). In addition, an isochronous readout of the 3x7 position sensor channels at a suitable sample frequency can be realized. The sensor positions are 1x sternum, 2x cardiac apex – medio-clavicular (5. ICR), 1x torso, lateral le., 1x spine (7.-8. th), 1x A. temporalis le., and 1x A. radialis left-hand side. The aim of the used two-digit kilohertz range including oversampling is to ensure data robustness. A temporal shift of only 20 ns ensures real-time measurement of ballistocardiographic parameters.

The measurements are carried out simultaneously with the recordings of the general patient assessment. Participation in the study will not give subjects any further inconvenience with regards to additional assessment procedures (e.g., 6-minute-walking- test) or separate visits to the measurement environment. All conducted tests take place as part of the diagnostic-therapeutic routine. The only change in the procedure concerns the additional equipment with the measuring sensors.

The subject is initially in a stationary position during the measurement (for about one minute). Further, a six minutes gait phase, as is commonly done within the general clinical routine assessment, will follow. After this physical stress on the subjects, there is a five-minute recovery period. Conversations should be avoided during the measurement.

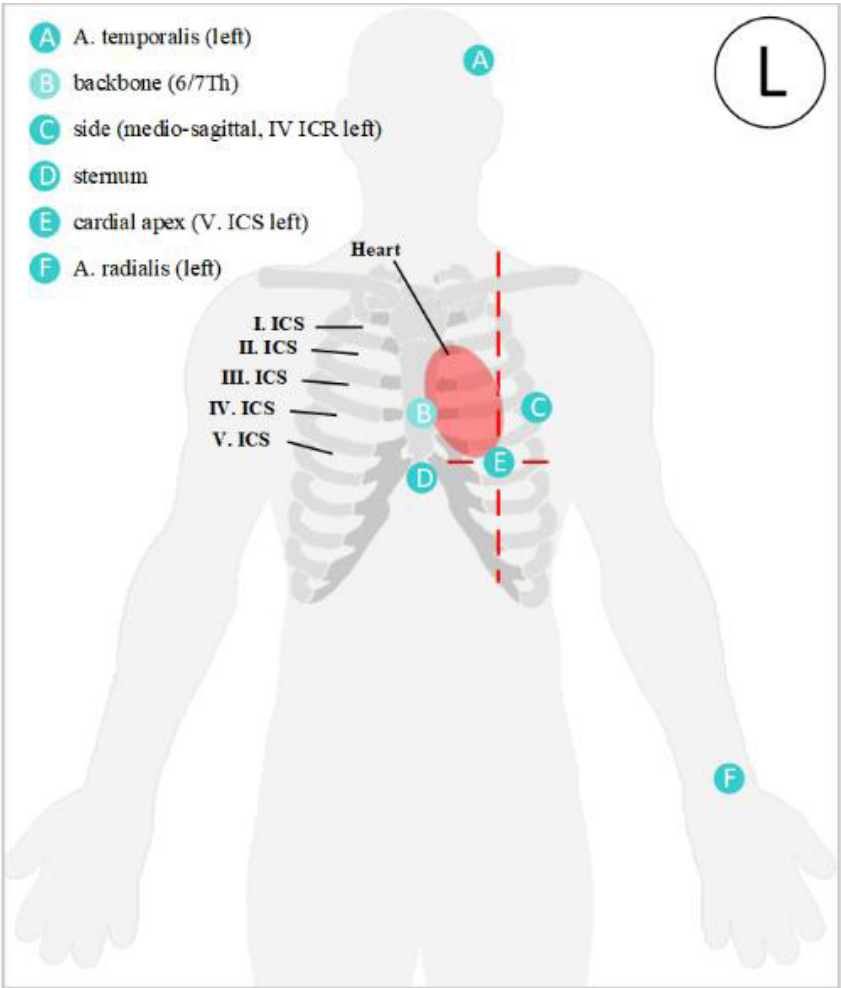


Figure 1. Model of the measure points.

The additional BCG sensor on the apex of the heart serves as a supplement to the first sensor since the electrode of the ECG is also attached to this position. This ensures a good recording of the acceleration at this.

Based on the hypothesis that there will be no subjects within the study population whose BCG/SCG signals differ fundamentally from the known morphology and the assumption that a homogeneous group exists under the given conditions more than 60 percent of the subjects have comparable features in the BCG or SCG signal, a rather small panel of subjects provides sufficient security for verification. In order to test the results of the difference between compensated and decompensated cardiac insufficiency subjects found in the literature see [10], 52 probands are needed with an effect size of 0.35. These subjects should be distributed as equally as possible to the four NYHA classes. In addition, the number of compensated and decompensated subjects should also be equally distributed.

3. Prospective Results

The main results of this study will be the validation of the data and results from the exploratory study with heart failure patients [10]. Hopefully, we are at least able to reproduce these results and give an overview of the performance of ballistocardiographic sensors in the diagnostics of cardiac patients. Moreover, we are confident to collect BCG or SCG signal data in the best quality state, so we can provide them to the scientific community soon as open data like it is requested by FAIR-principle. Those results will provide highly synchronous signal data with the best sample frequency we can reach (about 17 kHz) and the optimal entropy about the heart's movements and functions.

Another result of the performed measurements will be a deeper knowledge of the usability of specific digital triaxle acceleration sensors and other system specifics, like the clock frequency, to gain an optimal measurement result and furthermore, more knowledge about the measurement positions on a patient's body.

In the end, the ideal setup for those measurements is going to be presented. Moreover, the needed features for early detection of heart failure and the characteristics of the disease in BCG data will be shown and also the usability of this type of diagnostic for cardiac patients.

References

- [1] M.C. Wolf, N. Jähne-Raden, H. Gütschleg, U. Kulau, M. Kallenbach, and K.-H. Wolf, First Feasibility Analysis of Ballistocardiography on a Passenger Flight. *Stud Health Technol Inform* **264** (2019), 1648–1649.
- [2] A.O. Bicen, D.C. Whittingslow, and O.T. Inan, Template-Based Statistical Modeling and Synthesis for Noise Analysis of Ballistocardiogram Signals: A Cycle-Averaged Approach. *IEEE J Biomed Health Inform* **23** (2019), 1516–1525.
- [3] R. Joshi, B.L. Bierling, X. Long, J. Weijers, L. Feijs, C. van Pul, and P. Andriessen, A Ballistographic Approach for Continuous and Non-Obtrusive Monitoring of Movement in Neonates. *IEEE J Transl Eng Health Med* **6** (2018), 2700809.
- [4] J.M. Zanetti and K. Tavakolian, Seismocardiography: past, present and future. *Conf Proc IEEE Eng Med Biol Soc* **2013** (2013), 7004–7007.
- [5] O.T. Inan, P.-F. Migeotte, K.-S. Park, M. Etemadi, K. Tavakolian, R. Casanella, J. Zanetti, J. Tank, Funtova, G.K. Prisk, and M. Di Rienzo, Ballistocardiography and seismocardiography: a review of recent advances. *IEEE J Biomed Health Inform* **19** (2015), 1414–1427.
- [6] N. Jähne-Raden, K.-H. Wolf, and M. Marschollek, Signal Detection Accuracy of Digital Accelerometers for Ballistocardiographic Propose. In *2017 Computing in Cardiology Conference (CinC) Computing in Cardiology*, 2017

- [7] N. Nguyen, N. Jähne-Raden, U. Kulau, and S. Sigg, Representation Learning for Sensor-based Device Pairing. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* IEEE, 19.03.2018 - 23.03.2018, pp. 508–511.
- [8] J.H. Migueles, C. Cadenas-Sanchez, U. Ekelund, C. Delisle Nyström, J. Mora-Gonzalez, M. Löf, I. Labayen, J.R. Ruiz, and F.B. Ortega, Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Med* **47** (2017), 1821–1845.
- [9] N. Jähne-Raden, M. Marschollek, U. Kulau, and L. Wolf, HeartBeat the Odds. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems - SenSys '17*, R. Eskicioglu, ed. ACM Press, New York, New York, USA, 2017, pp. 1–2.
- [10] O.T. Inan, M. Baran Pouyan, A.Q. Javaid, S. Dowling, M. Etemadi, A. Dorier, J.A. Heller, A.O. Bicen, S. Roy, T. de Marco, and L. Klein, Novel Wearable Seismocardiography and Machine Learning Algorithms Can Assess Clinical Status of Heart Failure Patients. *Circ Heart Fail* **11** (2018), e004313.
- [11] N. Jähne-Raden, H. Gutschleg, M. Kallenbach, T. Feuerle, and U. Kulau, Poster: SCARAB² - Scalable, Robust and Adaptive on Board Ballistocardiography. In *2018 14th International Conference on Distributed Computing in Sensor Systems (DCOSS)* IEEE, 18.06.2018 - 20.06.2018, pp. 111–112.
- [12] N. Jähne-Raden, H. Gutschleg, M. Kallenbach, U. Kulau, and K.-H. Wolf, *Ballistogardiography on aircrafts*, German Medical Science GMS Publishing House, 2018.
- [13] M. Di Rienzo, P. Meriggi, E. Vaini, P. Castiglioni, and F. Rizzo, 24h seismocardiogram monitoring in ambulant subjects. *Conf Proc IEEE Eng Med Biol Soc* **2012** (2012), 5050–5053.

Addressing the Search Challenges of Precision Medicine with Information Retrieval Systems and Physician Readers

Kate Fultz HOLLIS^{a,1}, Kirk ROBERTS^b, Steven BEDRICK^a, William R. HERSH^a

^a*Oregon Health & Science University Department of Medical Informatics and Clinical Epidemiology, Portland, OR, USA*

^b*The University of Texas Health Science Center, School of Biomedical Informatics, Houston, TX, USA*

Abstract. The Text REtrieval Conference (TREC), co-sponsored by the National Institute of Standards and Technology (NIST) in the US and US Department of Defense, was started in 1992. TREC's purpose is to support research within the information retrieval community by providing the infrastructure necessary for large-scale evaluation of text retrieval methodologies. In 2017, the TREC Precision Medicine (Roberts et al., 2017) track grew from the Clinical Decision Support track and focused on a narrower problem domain of precision oncology. After three years of computer runs being evaluated for relevance by physician readers, we provide a unique perspective of how to evaluate computer-generated articles and clinical trials pulled from PubMed and Clinicaltrials.gov to find relevant information on medical cases.

Keywords. Precision medicine, information retrieval

1. Introduction

Starting in 2017, the Text REtrieval Conference (TREC), co-sponsored by the US National Institute of Standards and Technology (NIST) and the US Department of Defense, produces as part of its large information retrieval research program a project devoted to information retrieval focused on precision medicine. The 2017 “Overview of the TREC 2017 Precision Medicine Track” reported on search challenges with “precision medicine.” Many more treatment options become available with information tailored to the individual patient. The abundance of clinical information can easily overwhelm clinicians attempting to stay up to date with the latest findings and can easily inhibit a clinician's attempts to determine the best possible treatment for a particular patient. As a forerunner to the Precision Medicine Track, three consecutive years of the TREC Clinical Decision Support (CDS) track sought to evaluate information retrieval (IR) systems that provide medical evidence to the point-of-care. The CDS track then focused on the needs of precision medicine so IR systems can focus on this important issue [1].

The years 2017, 2018, and 2019 Precision Medicine Tracks focused on a single field, oncology, for a specific use case: genetic mutations of cancer. The main idea behind

¹ Corresponding Author, Kate Fultz Hollis; E-mail: fultzhol@ohsu.edu.

precision medicine is to use detailed patient information (largely genetic information in most current research) to identify the most effective treatments. Improving patient care in precision oncology then requires both (a) a mechanism to locate the latest research relevant to a patient, and (b) a fallback mechanism to locate the most relevant clinical trials when the latest techniques prove ineffective for a patient. In the first part, the track continues the previous Clinical Decision Support track [2], while in the second part the study expands the task to cover a new type of data (clinical trial descriptions). All of the TREC Precision Medicine tracks have used synthetic data to create topics of clinical cases for the computer search. All topics were created synthetically because it was difficult to obtain actual patient data for the project [3].

Detailed results from the TREC Precision Medicine Tracks are available in the overviews published on the TREC website. This paper focuses on the process of recruiting physicians to read the output from the computer runs, analyze the relevancy of the article or clinical trial for information about the topic, and help researchers provide information about the accuracy of the computer to find useful information for the physician.

2. Methods

A pooling strategy was used to determine the set of articles and trials to be judged. Slightly different pooling strategies were used for each year for each type of result, but the most common method is as follows. All articles and trials that were within the top 10 results for any research participant (i.e., the research groups searching for the article and clinical trial documents) were included in the pool. Additionally, a 10% sample of the results in position 11 to 100 were added to the pool. The pooled articles and clinical trials for each topic were judged by physicians who, in most cases, were graduates of the Oregon Health & Science University (OHSU) biomedical informatics graduate program. In the case of the clinical support track where physicians evaluated whether a clinical visit was relevant to the case, the challenges included terminology differences, negation, and time aspects having the most impact on retrieval inaccuracy. Edinger et al. describe in detail that these challenges also arise during cohort identification in medical text [4].

The cohort of graduates from OHSU ideally needed to have the following characteristics to become readers for the project: a certificate or master’s degree in biomedical informatics, a medical or doctor of osteopathic medicine degree to evaluate relevancy appropriate in a clinical situation, and an ability to quickly review many documents from a website created for the project. The number of topics (topic examples are in Table 1), documents found (total number of abstracts from PubMed and clinical trials from ClinicalTrials.gov), and readers for each year in the precision medicine track are outlined in Table 2.

Table 1. Example topics (3)

Disease	Variant	Demographic
melanoma	BRAF (V600E)	64-year-old male
melanoma	high serum LDH levels	69-year-old female
medullary thyroid cancer	RET	45-year-old female

Table 2. Number of topics, documents and readers

	2017	2018	2019
Number of Topics	30	50	40
Total Number of Documents	36081	36617	32162
Number of Physician Readers	18	26	19

2.1. Recruitment

Over three years of reader recruitment for the Precision Medicine track, we had to slightly change where we got readers for the project. In all three years, the chair of the OHSU Department of Medical Informatics and Clinical Epidemiology (DMICE) would send a letter to DMICE alumni roughly three months before the October 1 deadline. As the numbers in Table 2 indicate, we had many more topics in 2018 and so also sent the letter to chair people in other departments of biomedical informatics for their suggestions of clinical informatics fellows and doctorate students who might be interested in being readers for the project.

The incentives for readers to work on the project were to learn more about information retrieval and receive payment for reading. Noted in our recruitment letter for the work, we paid a set amount per hour. We paid for quantity of articles judged, not by time, but experience has shown that almost everyone can average one minute or less per article.

3. Results and discussion

The primary purpose for the physician readers in this project was to determine if each article and clinical trial found for a particular topic was 1) human medicine, 2) disease is relevant, partially relevant, or not relevant, 3) whether gene variants were discussed, and 4) whether the demographics (age and sex) matched (see figure 1 for screenshot of OHSU website for viewing articles and trials the computer selected). Details on the relevancy statistics for TREC Precision Medicine Tracks are in (1) and (2) and are forthcoming in 2019. The input of the physician readers has helped the project in many ways outlined in the following three years.



Figure 1. Example of OHSU website for reading articles captured in the computer runs.

3.1. 2017

The recruitment for readers began on August 17, 2017, and 18 readers finished 18 topics in 30 days. The remaining 12 topics were finished by October 8. In 2017, there were roughly 9% judged definitely relevant out of 22,642 literature articles, and 3% judged definitely relevant out of 13,441 clinical trials. We had and continue to have an international contingent of physician readers with many throughout the US and some in South America, Taiwan, and the Middle East. Readers mentioned that it was particularly difficult to find genes mentioned in the articles and usually would need to search in the full text of the article (the web interface we use shows article abstracts only). In terms of the web interface to read the abstracts and clinical trials, readers had, in some cases skipped articles and trials and had difficulty finding the particular missing article or trial in the web interface. Some of the time involved to get topics finished was due to skipped articles. By 2019, the website had been refined so that missing articles while working on the web interface was not a problem.

3.2. 2018

The second year of the Precision Medicine track included almost 50 topics, and it was necessary to recruit outside of the OHSU graduate program and in particular we found doctorate candidates in National Library of Medicine programs across the United States. The recruitment for readers began on August 22, 2018, and on average, 26 readers finished 25 topics in 21 days. The remaining 25 topics were finished by October 1. In 2018, there were roughly 15% judged definitely relevant out of 22,429 literature articles, and 6% judged definitely relevant out of 14,188 clinical trials. Readers from the doctorate programs would, in some cases, complete ranking topics faster than the physician readers and in other cases had the most questions and took more time. While we felt it was important to include informatics doctorate students without medical degrees, the physician readers had more skill in identifying articles and trials that would help for a particular topic.

3.3. 2019

At the time of this writing, the Precision Medicine Track reader analysis is still in operation (results due October 1, 2019). The recruitment for readers began on August 12, 2018, and on average 19 readers finished 19 topics in 16 days. The website took longer this year to produce, and there were more questions about how to match what the computer found as a genetic variant in articles and clinical trials.

The challenge with working with physician readers is that most have full-time clinical responsibilities (one reader was on call for forest fire duty in California in addition to clinical duties), but all readers were able to finish on time and in the case of two readers, one reading up to 1212 combined abstracts and trials for one topic in 10 days and another did four topics in 2 weeks. 15 readers in 2019 have made all of the previous Precision Medicine Tracks and look forward to working with us again. Many of the readers provide us with questions to help make the search for precision medicine topics better: e.g., the computer runs were sometimes not bringing in any relevant articles and how do the research groups change the algorithm to bring in more relevant articles

and trials. Some readers tried to do the reading and analysis on mobile devices. That is not possible now but could be considered in the future to enable more readers to work with the TREC website from OHSU DMICE.

It is possible that duplicate judging might improve the process as readers thought that in some cases, they were not as familiar with the case and suggested that another expert review their cases. We used to do duplicate judging (in the CDS Track) with kappa analysis, but the results always seemed to be about the same (good but not great consistency). While readers noted that there seemed to them poor relevancy of the articles and clinical trials chosen by computer, there appear to be improvements from 2017 to 2018, and there were fewer comments by readers about how few relevant articles there were in 2019 (only one reader and he was new).

The TREC assessment exercise can improve how we look for cases that closely resemble a clinical problem in precision medicine. The data from the physician readers has allowed more work on computer analysis of medical cases and eventually help with better searching for clinical cases in electronic health records.

4. Acknowledgements

KFH, KR, SB, and WTH and other organizers are grateful to the National Institute of Standards and Technology (NIST) for funding the assessment process for TREC.

References

- [1] K. Roberts, D. Demner-Fushman, E.M. Voorhes, W.R. Hersh, A.J. Lazar, Overview of the TREC 2017 Precision Medicine Track, <https://trec.nist.gov/pubs/trec26/papers/Overview-PM.pdf>, *Text Retr Conf.* (2017), 1–6.
- [2] K. Roberts, D. Demner-fushman, E.M. Voorhees, W.R. Hersh, Overview of the TREC 2016 Clinical Decision Support Track, <https://trec.nist.gov/pubs/trec25/papers/Overview-CL.pdf>, *Text Retr Conf.* **2** (2016), 1–14.
- [3] K. Roberts, D. Demner-Fushman, E.M. Voorhes, W.R. Hersh, A.J. Lazar, Overview of the TREC 2018 Precision Medicine Track, <https://trec.nist.gov/pubs/trec27/papers/Overview-PM.pdf>, *Text Retr Conf.* (2018), 1–6.
- [4] T. Edinger, A.M. Cohen, S. Bedrick, K. Ambert, W. Hersh, Barriers to retrieving patient information from electronic health record data: failure analysis from the TREC Medical Records Track, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540501/pdf/amia_2012_symp_0180.pdf, *AMIA Annu Symp Proc.*, 2012.

Argentinian Digital Health Strategy

Daniel A. RIZZATO LEDE^a, Federico A. PEDERNERA^a, Emiliano LÓPEZ^a,
Cintia D SPERANZA^a, Carlos GUEVEL^a, Jesse J. MAID^a, Patricia MAC CULLOCH^b,
Florencia ROLANDI^b, Fabiana AYALA^b, Daniel A. ABADIE^b, María I. BAQUÉ^b,
Fernando GASSINO^{a,d}, Fernando CAMPOS^{c,d}, Diego KAMINKER^c, Cintia A. CEJAS^a,
Alejandro LÓPEZ OSORNIO^a and Adolfo RUBINSTEIN^a

^aMinistry of Health, Argentina

^bMinistry of Modernisation, Argentina

^cHL7 Argentina

^dHospital Italiano de Buenos Aires, Argentina

Abstract. Digital Health is one of the three pillars for the effective implementation of Universal Health Coverage in Argentina. The Ministry of Health published the National Digital Health Strategy 2018-2024 in order to establish the conceptual guidelines for the design and development of interoperable health information systems as a state policy. The World Health Organization "National eHealth Strategy Toolkit", "Global Strategy on Digital Health" and other international and local evidence and expert recommendations were taken into account. The path to better healthcare involves adopting systems at the point of care, allowing for the primary recording of information and enabling information exchange through real interoperability. In that way, people, technology and processes will synergize to enhance integrated health service networks. In this paper, we describe the plan and the first two years of implementation of the strategy.

Keywords. eHealth, electronic health records, health information exchange, national health programs, health level seven, SNOMED.

1. Introduction

The United Nations have set universal healthcare coverage (UHC) as a target for sustainable development goal 3.8, to be achieved by 2030[1].

Only the appropriate use of digital technologies can enable the development of sustainable health systems that promote universal access to care. To deliver this potential, digital health initiatives must be guided by a robust strategy that integrates financial, organizational, human and technological resources [2]. History shows that ill-coordinated initiatives lead to vertical or stand-alone ICT solutions, that although well intended, often result in information fragmentation, and consequently poor delivery of care. Additionally, developing countries face specific challenges such as poor economics, political uncertainty, and the lack of cutting-edge infrastructure, that hampered the quality of healthcare [3].

Argentina has a complex federal healthcare system, where the national government, 24 jurisdictions and three subsystems interact: Public care, social security and private insurances. Many of its shortcomings come from its segmented and highly fragmented system, which also translates into the generation of information siloes. In many health centres in the country, health records are still kept in manual registries and logbooks. In a preliminary survey, at least 80% of the public healthcare facilities do not have

computers or Internet connection. This often results in duplication, over-reporting and underreporting for public health. To undertake this situation, in 2018 the Ministry of Health (MoH) promoted a National Digital Health Strategy, rising capacity building, engaging multiple stakeholders and promoting the development and interconnection of interoperable information systems, based on standards. This paper reflects the foundations and steps taken during the last two years to advance this mid and long term strategy.

2. Materials and Methods

2.1. Setting

Argentina is a Federal Republic subdivided into 24 jurisdictions, with an area of more than 1 million square miles and located in the southern half of South America. Its population is around 45 million people, unevenly distributed mainly in urban and suburban areas. Beyond the recurrent economic crises, it is an upper-middle-income country and has a developed healthcare system. However, health outcomes lag behind the country's potential, considering spending per capita of nearly 10 % of its gross domestic product. We have a mixed healthcare system with three components: public care, social security and private insurances. The social security subsystem is based on job taxes and covers 60% of the population, including active/retired workers and their families. Almost 2 million people (4%) pays for private insurance, leaving 36% (16 million people) with exclusive public care coverage based on jurisdictional states resources. There are significative gaps in morbidity and mortality outcomes between regions and subsystems. Since 2018, the National MoH has embraced an action plan to implement public health reform. Overall, the ultimate goal was to provide actual UHC by strengthening provincial health systems [4].

The existing fragmentation and segmentation of the healthcare system hamper the aggregation of patient data. Even though there is a need for digitalization, the creation of a single national health database is not possible given the political and health model. The incipient and progressive individual computerization of each system does not spontaneously solve this difficulty. Therefore, a distributed but interconnected health information ecosystem was selected as a feasible approach.

2.2. Methods

Our National Digital Health Strategy was built, taking into account international recommendations, like the World Health Organization toolkit for National eHealth Strategies, and their recent global strategy on digital health. We also participated in working groups such as HL7 International, SNOMED International and the Global Digital Health Partnership (GDHP). We capitalized regional initiatives like Information Systems for Health (IS4H) and RELACIS from the Pan American Health Organization (PAHO)[5], and Inter American Development Bank sponsored RACSEL. We interacted with Latin American leaders in charge of other national digital transformation experiences like Uruguayan SALUD.UY. Our approach is also grounded in the Principles for Digital Development and the subsequent Principles of Donor Alignment for Digital Health. These principles highlight the importance of supporting strong

national digital health strategies for sustainable country ownership, avoiding project overlap and fostering coordination and alignment of implementation activities [6].

However, it was essential to uphold the local existing resources and capacities, based on public, private and non-government organizations stakeholders who have been using Information Technologies for health services improvement for many years [7]. Moreover, the OECD has highlighted the ambitious general digital agenda and the efforts taking place in Argentina to digitalize and improve data governance in its public sector and build the foundations for a digital government [8].

3. Results

Through Resolution 189/2018, the highest health authority in Argentina settled the way forward to build, in collaboration with provincial governments and referents of the three subsystems, a network of information systems focused both on the population and individual health of each patient. The MoH created a National Directorate of Health Information System to coordinate such a strategy. It was essential to work on governance, making agreements with each province within the UHC framework plan.

The investments that each jurisdiction makes in its digitization projects are reinforced with national and international funds. Multilateral funding organizations (such as the World Bank and Inter American Development Bank) support projects like so-called SUMAR, which strengthen public health coverage to protect the most vulnerable populations [9]. Recently, they have included the Digital Health Strategy goals and will give economic incentives to the provinces that accomplish them for the next four years. As a result, every digital health transformation project should be aligned with the National Strategy. Additionally, the MoH got the funding for a National Connectivity Plan for public healthcare facilities. It will reach 2,000 remote primary care centres in the next two years, using satellite connections and optic fibre.

The Digital Health Strategy has two consecutive phases. Phase 1: Consensus, Infrastructure and Scalable Projects (2018-2019), which define the technical, regulatory and political foundations, with the participation of the community of experts and jurisdictions. Phase 2: Deepening, scaling up, and maturation (2020-2024) focuses on the adoption extension of the Digital Health tools and the functional enhancement of Information Systems. The priorities for the first stage were to:

Define interoperability standards: SNOMED, FHIR, ICD, etc.

Make the Central Infrastructure available: the enterprise service bus (EBS).

Identify and integrate the information needs: governance on the data flow.

Enhance local information systems: licensed, homegrown, or open-sourced.

Define the architecture to share clinical records, with HL7-FHIR.

Inform and empower the patient through a Citizen Health Portal (*Mi Argentina*).

Advance in the legal framework for digital health transformation.

Grow capacity building: train human resources in health information systems.

Argentina joined the SNOMED International Consortium in 2018. The National Release Centre has published three local extensions during the last 18 months, including medication codes to allow interoperable ePrescriptions. The MoH stated SNOMED as the clinical terminology standard for electronic health record (EHR) systems, leaving ICD for statistical and other administrative tasks. In addition, FHIR (Fast Healthcare

Interoperability Resources) [10] a next-generation standards framework created by HL7, was chosen as the standard for health care data exchange.

The National Digital Health Network was launched in April 2019, using an enterprise service bus technological infrastructure to connect different domains (EHR systems) across the country, as shown in Figure 1. We defined the required interoperability standards based on international recommendations and extensive local expertise in the field. By involving multiple stakeholders, we understood their views and gained their collaboration, support and endorsement of the strategy. A joint team including MoH, HL7 Argentina and Hospital Italiano (HIBA) experts developed the HL7-FHIR Enterprise Service Bus. The network enables the communication between domains but does not change the current processes of storage of health information. The data remains in each information system repository, and is only communicated to other domains when necessary (i.e. when the patient visit another clinic).

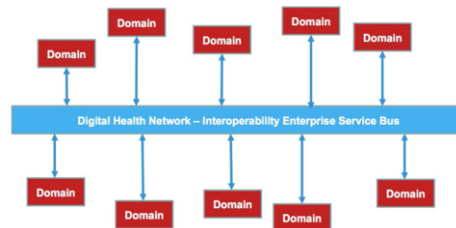


Figure 1. Simplified ESB architecture of the National Digital Health Network.

The MoH acquired RedHat OpenShift® licenses and services to support such infrastructure, hosted at the main Government Datacenter (ARSAT). The MoH team has been adopting this open-source container application platform based on the Kubernetes container orchestrator for enterprise application development and deployment, to guarantee the Health Network required availability and performance level.

After two years, every province (24) has defined its Digital Agenda for Health. Regarding their information systems, they have freely decided whether to acquire commercial EHR licenses, develop locally or adopt open-source software. There are advanced Projects in 16 of 24 states, where they selected pilot areas to carry out the digital transformation. Human resources, technology and processes were aligned to fulfil the project goals. Adopting a change management strategy was crucial for EHR implementations. The national MoH helped local teams to implement standards like SNOMED and FHIR in their systems. Vendors are progressively adopting standards to comply with National recommendations. Up to date, there are more than 2 million patients registered in the Network: it means that they have received care in an institution with interoperable EHR, and if they attend to another clinic, it would be possible to ask individually for such records through the network.

In collaboration with the Ministry of Modernisation, we launched a Citizen Health Portal. As of October 2019, more than 2 million users had been registered on the digital public service delivery platform *Mi Argentina*. It gives personal information like the Digital Vaccination Card, but also works as a patient privacy manager, to configure the opt-out consent for the information exchange trough the EBS health network.

We developed and implemented a National Telehealth Platform to perform eReferrals and eConsultations nationwide. It allows every clinician to interact with other healthcare professionals (specialists) using a personal computer or a mobile device in a secure environment, even if they do not have an EHR in place. As the Telehealth

Platform is connected to the National Enterprise Service Bus, different EHR can interoperate with it.

Besides the existing legal framework that rules Health IT implementations (Digital Signature, Personal Data Protection, EHR, etc.), the MoH dictated several resolutions to promote digitization while caring for privacy, confidentiality and security. In addition, there is a Digital Health Law project to be discussed at the Congress. Every jurisdiction can dictate their legal regulations, or endorse national laws.

We promoted several initiatives to enhance capacity building. In 2018 the MoH, PAHO and HIBA launched the “100 Leaders” Training Plan, using an online AMIA 10x10 course. There is also a monthly open course on standards and Conectathon, both onsite and online. We developed a basic online course on EHR fundamentals for more than 1,500 healthcare staff. The next big step is introducing these topics in healthcare professionals’ undergraduate curricula.

4. Conclusion

The Digital Health Strategy is not an end in itself but has the vision of reducing quality gaps in health care, implementing information systems that identify the characteristics and needs of the population, allowing longitudinal and comprehensive monitoring of people throughout the entire health system and providing innovative tools to health professionals and patients.

This paper underlines the role of the MoH in setting a National Digital Health Strategy for Argentina. Even though it has been a short period, since 2018, there have been several achievements that should be sustained. Besides its mid and long term planning, the strategy itself has to prove that it can succeed the time challenge of political transitions. Using international benchmarks and strategies like IS4H could be useful to support this initiative.

References

- [1] SDG 2030, (n.d.), <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>.
- [2] *WHO guideline Recommendations on Digital Interventions for Health System Strengthening*, WHO, Geneva, 2019.
- [3] D. Luna, A. Almerares, J.C. Mayan, F. Quirós, and C. Otero, Health Informatics in Developing Countries: Going beyond Pilot Practices to Sustainable Implementations: A Review of the Current Challenges, *Healthc Inform Res.* **20** (2014), 3–10.
- [4] A. Rubinstein, M.C. Zerbino, C. Cejas, and A. López, Making Universal Health Care Effective in Argentina: A Blueprint for Reform, *Heal. Syst. Reform.* **4** (2018), 203–213.
- [5] *Information Systems for Health - IS4H*, (n.d.), <http://www.paho.org/ish/index.php/en/>, PAHO
- [6] A. Labrique, L. Vasudevan, G. Mehl, E. Rosskam, and A.A. Hyder, Digital Health and Health Systems of the Future., *Glob. Heal. Sci. Pract.* **6** (2018), S1–S4.
- [7] D. Luna, C. Otero, F. Plazzotta, F. Campos, and F. Quirós, *Sistemas de Información para la Salud (Spanish Edition)*, 1st ed., Hospital Italiano de Buenos Aires, Buenos Aires, 2018.
- [8] *Digital Government Review of Argentina: Accelerating the Digitalisation of the Public Sector*, <https://doi.org/10.1787/354732cc-en>, OECD, Paris, 2018.
- [9] *WB Supports Expansion of Public Health to Benefit 15M Argentines*, <https://www.worldbank.org/en/news/press-release/2018/05/21/proyecto-salud-publica-argentina>., World Bank, 2018.
- [10] M. Lehne, S. Luijten, P. Vom Felde Genannt Imbusch, and S. Thun, The Use of FHIR in Digital Health - A Review of the Scientific Literature., *Stud. Health Technol. Inform.* **267** (2019), 52–58.

Assessing Opioid Use Patient Representations and Subtypes

Sarah MULLIN^{a,1} and Peter ELKIN^a

^a *University at Buffalo, Buffalo, NY, USA*

Abstract. Precision medicine, diagnosis, treatment, and prevention that accounts for natural human variability, can be beneficial for complex populations. The opioid user population is heterogeneous, characterized by many disorders, medications, and procedures. Using Electronic Health Record data, we create a patient representation, finding similarities between structured data, and then cluster the patients into patient subtypes. These subtypes can then be used for subsequent analysis.

Keywords. electronic health records, precision medicine, deep learning

1. Introduction

Health professionals are required to make analytic decisions influenced jointly by evidence-based medicine, prior practice knowledge, and specific patient attributes. Precision medicine aims to find appropriate treatments and diagnoses based on an individual patient since no method is ideal for every patient. Opioid-related adverse events and addiction are the main public health concern in the United States. Since 1999, more than 700,000 people have died from a drug overdose [8]. Partially due to over-prescribing and increased availability of potent synthetic opioids such as fentanyl, the opioid crisis has brought to light the need for finding better treatment options that are tailored to the patient for pain and addiction [7, 13].

Previous literature has demonstrated that high opioid dosage, concurrent use of sedative-hypnotics, substance abuse, and mental health comorbidities are associated with an increased risk of overdose [9]. Very few models have assessed the entire patient clinical picture and those that do focus mainly on medication data, require reliance on some sort of manual feature creation, and specifically are tasked to predict overdose [1; 5]. Therefore, characterizing an opioid prescribed population for specific pain treatment candidates, adverse opioid drug-drug interactions, and unknown risk factors for opioid overdose could be beneficial.

As a step toward precision medicine for an opioid prescribed population, using electronic health record data (EHR) provides the benefit of a large source of longitudinal data that encompasses the clinical picture with diagnoses, medications, and procedures. Retrospective observational data is especially useful for discovering unknown associations and can give a basis for finding similar patients [11]. However, EHR data suffers from an abundant amount of sparse attributes in the form of structured terminologically codified data and unstructured clinical text [10].

¹ Sarah Mullin, E-mail: sarahmul@buffalo.edu.

Deep learning advances have made it possible to incorporate high-dimensional data as input, as opposed to a select set of manually engineered features, by learning low dimensional dense embeddings [2, 6, 12]. Patient representations can be learned from neural networks that encode meaning and relationships into the dimensions. These patient vectors have been shown to increase predictive performance, assess patient similarity, and create patient subtypes [3, 14].

Using a paragraph vector model, we learn low dimensional dense patient representations from codified diagnoses, procedures, and medications for an opioid prescription population. To analyze these representations, we look at how well the vectors cluster patients into opioid user subtypes and known opioid misuse risk factors.

2. Methods

Two retrospective EHR databases, an outpatient practice and a hospital, were collected from January 2013 to December 2017 for patients who received an opioid prescription within the time frame. The data was approved as exempt research by the University at Buffalo IRB. Merger and de-identification of the data with assigned unique research identifiers were done with patient matching on social security number, date of birth, last name, and first three letters of the first name. Encounters for patients with a cancer diagnosis or those in palliative care were removed from the analysis. In order to assess the effects of opioid usage, patients were retained if they had at least two visits.

Due to the large volume of terminological codes and version differences across the time window, clinical classification software (CCS) was used to classify ICD-9 and ICD-10 diagnoses and current procedural terminology (CPT) codes to upper-level classes. Medications were mapped from national drug codes and local facility drug codes to the drug classes opioids, muscle relaxants, benzodiazepines, zolpidem, and Schedule II stimulants within the 2018 IBM® MarketScan® RED BOOK Database (www.ibm.com/Watson-Health/MarketScan). Finally, additional demographic variables, age at first opioid prescription, gender, race, ethnicity, and marital status were included. Race and ethnicity were standardized according to the Agency for Healthcare Research and Quality (AHRQ). Codes for drug overdose and adverse events, defined by ICD and manual review of clinical text, were removed from the patient representation for downstream prediction tasks.

To build the distributed representations of patients and their specific medical codes, we implemented the distributed bag of words algorithm (PV-DBOW) with 200-dimensional feature vectors and negative sampling [4]. Codes that were used less than two times in the analysis were removed. A patient vector consisted of all codes in temporal order by encounter date with randomization of medication, diagnosis, and procedure codes within an encounter. A temporal embedding was not considered due to the small patient sample in relation to deep learning methods. We trained the model for 500 epochs. To visualize our patient representations, we used t-distributed stochastic neighbour (t-SNE) embedding to collapse the representations into two dimensions. Cosine similarity, a simple mean of the projection weight vectors, is used to assess similar codes. Finally, we used k-means with Euclidean distance to create patient subtypes from the patient representations and assessed the distribution of features within those groupings.

3. Results

After removal of patients with cancer and palliative care diagnoses, the patient representation was trained on 3,986 patients. The features consisted of 129 distinct procedural CCS codes, 44 distinct medication codes, 256 distinct diagnostic CCS codes, and four demographic variables. For diagnostic codes, ‘residual codes; unclassified,’ ‘medical examination/evaluation,’ and ‘administrative/social admission’ were removed. 13% of patients had a diagnostic code for opioid dependence, with 11.4% on buprenorphine treatment and 4.1% on methadone treatment. Opioid-related adverse events and overdoses made up only 3.6% of cases. Mental health-related diagnoses were recorded in 60% of the population, and non-opioid substance abuse was recorded in 8.1% of the population. Analysing the demographic features of the cohort, 64% were female, 58.8% were White, 40.14% were Black, and 3.7% were Hispanic. On average, there were 29 codes per patient.

Cosine similarity was assessed between the codes in the embedding (Table 1). Opioid medication codes, psychological and mood disorder codes, and procedures relating to pain were similar, implying the associations of previously reported risk factors in an opioid-related population are being learned by the distributed representation.

Table 1. Validation examples of patient representation code embeddings measured by cosine similarity. (p) indicates CCS procedure code as opposed to diagnosis code.

Code	1 st Similar (cosine similarity)	2 nd Similar (cosine similarity)	3 rd Similar (cosine similarity)	4 th Similar (cosine similarity)
Methadone	CCS 211: Other connective tissue disease (0.59)	Morphine Long-Acting (0.57)	CCS 204: Other non-traumatic joint disorders (0.56)	Hydromorphone Short-Acting (0.53)
CCS 651: Anxiety disorders	CCS 661: Substance-related disorders (0.67)	CCS 663: History of mental health and substance abuse codes (0.65)	CCS 652: Malaise and fatigue (0.62)	CCS 657: Mood Disorders (0.61)
CCS(p) 155: Arthrocentesis	CCS(p) 173: Diagnostic procedures on skin and subcutaneous tissue (0.40)	CCS(p) 38: Diagnostic procedures on lung and bronchus (0.34)	Tramadol Long-Acting (0.27)	Orphenadrine (0.25)

Visualizing our patient representations, we see that opioid abuse, misuse, and poisoning, opioid dependence, mental illnesses, and other drug abuse are not well represented in two dimensions of the embeddings (Figure 1).

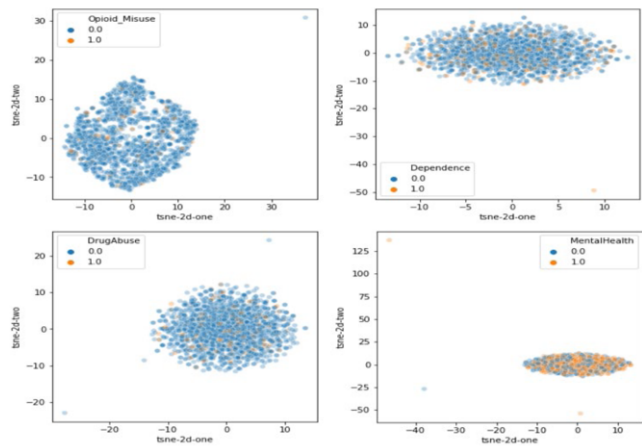


Figure 1. T-SNE patient representations of opioid-related outcomes misuse and dependence and other risk factors, drug abuse and mental health.

Using the elbow method and distortion, 8 clusters were found to be optimal for k-means clustering. Compressing the data into two dimensions using t-SNE, the clusters were plotted for visual inspection (Figure 2). The clusters have well-distinguished groupings. Cluster 6 is characterized by opioid dependence (58%) and opioid poisonings and misuse (44%). Cluster 0 has a high proportion of hydrocodone, fentanyl, and nutritional disorders. Cluster 1 has a high proportion of buprenorphine users, benzodiazepines, and mental health disorders, particularly substance-related disorders, anxiety, and mood disorders. Cluster 2 has a high proportion of opioid users related to bone and tissue diseases and spine and back pain. Cluster 3 is characterized by joint and nervous system disorders, muscle relaxers, and hydrocodone or buprenorphine use. Cluster 4 is also characterized by connective tissue and joint disorders, but patients are primarily on tramadol or hydrocodone. Cluster 5 is characterized by urinary infections and disorders and arthrocentesis. Cluster 7 is characterized by hypertension, diabetes, and gastrointestinal disorders.

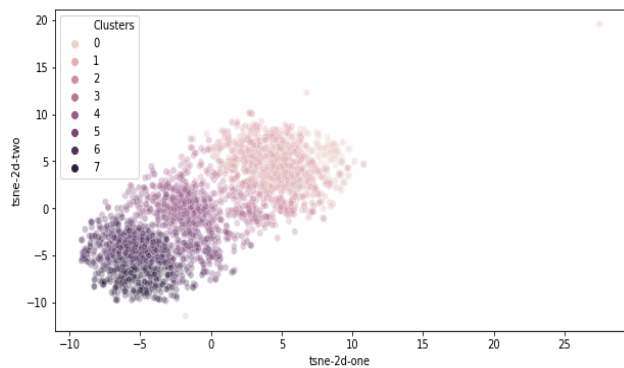


Figure 2. K-means clusters using t-SNE dimensions for opioid prescribed patient population.

4. Discussion and Conclusion

Patient subtyping, seeking similar groups of people in an opioid patient population with different diagnostic, medication, and procedural patterns, is important to address the heterogeneity in patients and to find potential new treatments or treatment plans. By learning a patient representation of EHR codes as opposed to primarily manually featured engineered features, we can learn new risk factors for predicting multiple tasks, including opioid misuse, in a general and overly complex cohort of opioid users.

Shown in the k-means clusters, patient representations can be grouped into subtypes that can be beneficial to individual treatment plans and policymakers for complex patient populations.

5. Acknowledgements.

This work has been supported in part by grants from NIH NLM T15LM012595, NIAA R21AA026954, and NCATS UL1TR001412. This study was funded in part by the Department of Veterans Affairs.

References

- [1] Z. Che, J. St Sauver, H. Liu, and Y. Liu, Deep Learning Solutions for Classifying Patients on Opioid Use, *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2018.
- [2] E. Choi, M.T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, Retain: An interpretable predictive model for healthcare using reverse time attention mechanism, *Advances in Neural Information Processing Systems*, 2016.
- [3] Y. Choi, C.Y.-I. Chiu, and D. Sontag, Learning low-dimensional representations of medical concepts, *AMIA Summits on Translational Science Proceedings*, 2016.
- [4] Q. Le and T. Mikolov, Distributed representations of sentences and documents, *International conference on machine learning*, 2014.
- [5] W.-H. Lo-Ciganic, J.L. Huang, H.H. Zhang, J.C. Weiss, Y. Wu, et al., Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions, *JAMA network open* **2** (2019), e190968-e190968.
- [6] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao, Kame: Knowledge-based attention model for diagnosis prediction in healthcare, *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, ACM, 2018.
- [7] M.A. Makary, H.N. Overton, and P. Wang, *Overprescribing is major contributor to opioid crisis*, British Medical Journal Publishing Group, 2017.
- [8] O. Overdose, *Understanding the epidemic*, Centers for disease control and prevention, 2017
- [9] T.W. Park, L.A. Lin, A. Hosanagar, A. Kogowski, K. Paige, and A.S. Bohnert, Understanding risk factors for opioid overdose in clinical populations to inform treatment and policy, *Journal of addiction medicine* **10** (2016), 369-381.
- [10] S.R. Raman, L.H. Curtis, R. Temple, T. Andersson, J. Ezekowitz, et al., Leveraging electronic health records for clinical research, *Am Heart J* **202** (2018), 13-19.
- [11] A. Sharafoddini, J.A. Dubin, and J. Lee, Patient Similarity in Prediction Models Based on Health Data: A Scoping Review, *JMIR medical informatics* **5** (2017), e7-e7.
- [12] Q. Suo, F. Ma, Y. Yuan, M. Huai, W. Zhong, et al., Personalized disease prediction using a CNN-based similarity learning method, *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2017.
- [13] N.D. Volkow and F.S. Collins, The role of science in addressing the opioid crisis, *New England Journal of Medicine* **377** (2017), 391-394.
- [14] J. Zhang, K. Kowsari, J.H. Harrison, J.M. Lobo, and L.E. Barnes, Patient2vec: A personalized interpretable deep representation of the longitudinal electronic health record, *IEEE Access* **6** (2018), 65333-65346.

Automated Engineering for Health Smart Homes: Find a Way in the Jungle of Assistance Systems

Bastian WOLLSCHLAEGER ^{a,1} and Klaus KABITZSCH ^a

^aChair of Technical Information Systems, Technische Universität Dresden, 01062 Dresden, Germany

Abstract. Health Smart Homes are a promising part of digital personalized health care. However, engineering of the underlying residential assistance systems is a complex process that needs to be supported by computer-based design tools. This paper presents a formal definition for the engineering task and proposes a methodology for model-driven design process automation, that is implemented in a web-based pilot application. Using this approach for design support, the challenge to efficiently compose personalized assistance systems for patients can be coped with in the future, which reduces a major barrier for smart home assistance applications.

Keywords. Ambient Assisted Living, Health Smart Home, Digital Personalized Health, Interoperability, Telemedicine, Telehealth

1. Introduction

As a promising part of digital personalized health care, *Health Smart Homes* (HSH) [1,2] combine concepts from Smart Home, Ambient Assisted Living (AAL), and Telemedicine [3], aiming to improve the quality of health care delivery [4,5]. Yet, a significant proliferation of HSH technology into real-world applications is still missing [6,7], largely due to ineffectiveness of system design [8]. The range of available assistance components is overwhelming and these components are not always compatible with each other [4,9]. Even professional planners are not able to chart the area of assistance solutions comprehensively, as they can only be familiar with a small number of components. Thus, engineers, patients and health care providers are tangled up in navigating the confusing “jungle” of assistance solutions. They require support for component selection and composition in order to tailor assistance systems to patient needs.

We envision an automated design approach that takes patient requirements and needs into account and identifies several suggestions that each fulfill the patient’s individual demand for technical support [9]. As a prerequisite for this vision, this paper presents a formal definition for the engineering task and proposes a methodology for model-driven design process automation.

¹Corresponding author, e-mail: bastian.wollschlaeger@tu-dresden.de.

The remainder of the paper is structured as follows: The next section infers requirements for a design methodology and discusses related work for HSH engineering. The proposed new methodology is presented in Section 3, after introducing a formal problem definition for the engineering task. Section 4 discusses the methodology based on a pilot application. Finally, Section 5 summarizes the paper and provides further research areas.

2. Related Work

A design methodology needs to meet the following requirements: *Req. A) Efficiency and Sustainability* allows the methodology to incorporate the re-use of existing components. *Req. B) Allowing for individual Customizability* states that fine-grained customization capabilities are required [1, 10]. Since manually exploring the design space is not feasible due to a large number of potential solutions, *Req. C) Capability of Automation* demands the automation of key processing steps of the design methodology. Considering the variety of possible solutions, determining the most suitable solution will involve multiple criteria (such as costs, installation and maintenance effort). Thus, *Req. D) Multiple Solutions* states that a design methodology needs to be able to offer a "design space" containing possible alternative designs [11].

Nowadays, HSH systems are often developed from scratch as monolithic systems (*one-off approach*) [6, 12], applying generic high-level methodologies such as model-driven approaches (MDA). While this allows for a precise customization, these approaches become inefficient [11] as they incur an high effort for system design. Alternatively, HSH are sometimes designed based on coarse-grain bundles of existing modular components (*bundle approach*) [10], in an attempt to achieve a rudimentary level of customization with lower design effort. However, given the diversity of assistance solutions [9], the bundle approach does not allow a sufficiently fine-grained customization. *Integration-focused approaches* rely on standardization of data exchange, integration frameworks or plug-and-play mechanisms when composing systems from components-off-the-shelf [6]. Yet, while this reduces interoperability issues [1], selection of components and communication planning still needs to be done manually.

As a consequence, there is a need for a new methodology that is capable of fulfilling *Req. A) to D)*, especially providing individual customization capabilities in combination with an efficient design approach.

3. Automated Engineering of Health Smart Homes

3.1. Formal Task Definition

The task in HSH engineering is to design a suitable assistance system by selecting and composing assistance components to *Candidate Solutions* HSH_{cand} in order to meet the set of requirements posed by the HSH occupant (*Condition Portfolio* HSH_{req}). The design process builds upon a *Knowledge Base* HSH_{KB} containing the vocabularies of user requirements, assistance functions and assistance components [13] as well as their interdependencies. This task description can be formalized as follows:

Definition 1 Knowledge Base: Let be: (i) \hat{R} the universal set of user requirements; (ii) \hat{F} the universal set of assistance functionality; (iii) \hat{C} the universal set of assistance components; (iv) \hat{T} the universal set of semantic types of the information exchangeable by assistance components; (v) $R_f \subseteq \hat{R}$ the set of user requirements satisfied by each assistance functionality $f \in \hat{F}$; (vi) $F_c \subseteq \hat{F}$ one set of realized assistance functions for each assistance component $c \in \hat{C}$; (vii) then the knowledge base for HSH design is defined as $HSH_{KB} = (\hat{R}, \hat{F}, \hat{C}, \hat{T}, \cup R_f, \cup F_c)$ ².

Definition 2 Condition Portfolio: The patient's condition portfolio is defined as a subset R of user requirements s. t. $HSH_{req} = R \subseteq \hat{R}$.

Definition 3 Candidate Solution: Let be: (i) C the multi set of selected assistance components, with $\forall c \in C : c \in \hat{C}$; (ii) B the multi set of bindings (i. e. communication relations) modeling the information exchange between assistance components, with each binding $b = (c_o, c_i, T) \in B$ defined by the tuple of the information output and input components c_o and c_i , and the set $T \subseteq \hat{T}$ of information types of the exchanged data; (iii) then a candidate solution of a HSH is defined as $HSH_{cand} = (C, B)$.

Definition 4 HSH Engineering Task: Given the input of: (i) HSH_{KB} a knowledge base for HSH engineering and (ii) HSH_{req} a specific patient condition portfolio, the goal of the HSH engineering task is to find the set \mathcal{C} of valid candidate solutions HSH_{cand} with validity of a candidate solution $HSH_{cand} = (C, B) \in \mathcal{C}$ meaning that all patient requirements are met by the candidate solution: $R \subseteq \cup_{c \in C} \cup_{f \in F_c} R_f$.

3.2. Engineering Methodology for Health Smart Homes

Following the MDA paradigm to abstract from technology-specific information, we propose to add an additional technology-independent intermediate layer. As can be seen from the equation in Def. 4, when determining the validity of the candidate solutions, an intermediate step is taken to relate assistance components C to the requirements R . In accordance to this observation, a suitable intermediate model layer consists of an information flow graph containing technology-independent assistance functions F :

Definition 5 HSH System Specification: Let be: (i) F the multi set of required assistance functionality, with $\forall f \in F : f \in \hat{F}$; (ii) L the multi set of links modeling the information exchange between assistance functions, with each link $l = (f_o, f_i, T) \in L$ defined by the tuple of the information output and input functions f_o and f_i , and the set $T \subseteq \hat{T}$ of abstract information types of the exchanged information; (iii) then a system specification for HSH is defined as $HSH_{spec} = (F, L)$.

With the introduction of the intermediate model layer HSH_{spec} , the design task can be split into *Step 1) Computation of HSH system specification*, which yields technology- and vendor-neutral formalized specifications for the HSH, and *Step 2) Materializing the specification with assistance components*, which in turn yields several design suggestions (Figure 1). The different design alternatives may now be inspected in order to determine the most promising design proposition. As this depends on planner's experience and

²For readability we do not write the index for unions/sums that are unambiguous, i. e. $\cup F_c = \cup_{c \in \hat{C}} F_c$.

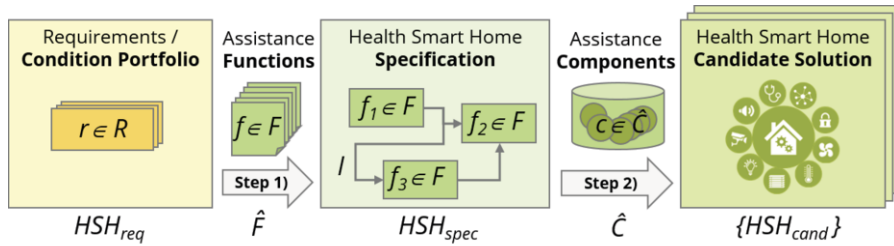


Figure 1. HSH engineering task and methodology of Automated Design for HSH

involves assessing additional criteria (e. g. costs) not modeled here and the decision may be rooted in the planner's experience, making the final decision is out of scope of this methodology and needs to be made in close collaboration with the patient.

4. Prototype and Discussion

To show the feasibility of the methodology, a pilot application has been implemented as a web-based suggestion system for AAL counselors, intended to facilitate the design and tailoring process of HSH solutions during patient consultation. After an interview-style analysis of user requirements, appropriate assistance functionality is determined. Finally, suitable assistance components are suggested to the AAL counselor and patient.

By composing existing assistance components based on their functionality, the presented methodology is able to efficiently provide several system proposals without the need for labor-intensively developing assistance systems from scratch. It therefore meets *Req. A) Efficiency and Sustainability*. Since the methodology is based on formal sets of user requirements, assistance functions and assistance component functionality, it can be supported by computer-based tools, fulfilling *Req. C) Capability of Automation*. The application of computer-based tools furthermore allows for an efficient exploration of the design space. The methodology is thus able to meet *Req. D) Multiple Solutions*. Finally, the input for the proposed design process is not confined to a fixed set of coarse patient types, but rather features a detailed vocabulary \hat{R} of possible user requirements and patient conditions, which in the pilot application have been identified according to personas and insights of health care providers. Thus, a fine-grained customization of the assistance systems is possible, fulfilling *Req. B) Allowing for individual Customizability*. Compared to the engineering methods *one-off* and *bundle* approach discussed in Section 2, the proposed methodology allows for an efficient exploration of the design space, thus lowering the overall costs of HSH engineering. Subsequently, *integration-focused* approaches can be applied for realization of the selected system design.

5. Conclusion

One major barrier for the proliferation of assistance systems in the context of smart homes is the tedious and ineffective design process. This paper proposed a model-based methodology for the automated engineering of HSH, enabling re-use of existing assistance components as well as fine-grained customization of the resulting assistance sys-

tems. By splitting the design task introducing a technology-neutral intermediate layer for system specification, the overall engineering task becomes more manageable and an abstraction from technology-specific information can be achieved.

Next steps include a further investigation into functional component models to ensure that the broad variety of real-world assistance functionality can be adequately modeled. Similarly, refining the universal set \hat{R} as well as optimizing the algorithm for design candidate identification are important tasks. Finally, tool-support for the actual process of functional component modeling shall reduce the impediments for component manufacturers to provide high-quality functional component models.

Acknowledgements

The work for this paper was funded by the European Social Fund and the Free State of Saxony (Grant no. 100310385). The authors would also like to offer special thanks to Tuan Linh Mai as well as Lena Otto for their valuable feedback on the manuscript.

References

- [1] A. J. Maeder and P. A. H. Williams, "Health smart homes: New challenges," *Studies in health technology and informatics*, vol. 245, pp. 166–169, 2017.
- [2] V. Rialle, F. Duchene, N. Noury, L. Bajolle, and J. Demongeot, "Health "smart" home: Information technology for patients at home," *Telemedicine Journal and e-Health*, vol. 8, no. 4, pp. 395–409, 2002.
- [3] L. Otto, L. Harst, H. Schlieter, B. Wollschlaeger, P. Richter, and P. Timpel, "Towards a unified understanding of ehealth and related terms – proposal of a consolidated terminological basis," in *Proceedings of the 11th BIOSTEC Conference - Volume 5: HEALTHINF*, INSTICC. SciTePress, 2018, pp. 533–539.
- [4] R. Haux, S. Koch, N. H. Lovell, M. Marschollek, N. Nakashima, and K.-H. Wolf, "Health-enabling and ambient assistive technologies: Past, present, future," *Yearb Med Inform*, pp. 76–91, May 2016.
- [5] C. Kruse, N. Krowski, B. Rodriguez, L. Tran, J. Vela, and M. Brooks, "Telehealth and patient satisfaction: a systematic review and narrative analysis," *BMJ Open*, vol. 7, pp. 1–13, 2017.
- [6] M. Memon, S. Wagner, C. Pedersen, F. Beevi, and F. Hansen, "Ambient Assisted Living Healthcare Frameworks, Platforms, Standards, and Quality Attributes," *Sensors*, vol. 14, no. 3, 2014.
- [7] L. Harst, P. Timpel, L. Otto, B. Wollschlaeger, P. Richter, and H. Schlieter, "Identifying obstacles and research gaps of telemedicine projects: Approach for a state-of-the-art analysis," *Studies in health technology and informatics*, vol. 247, pp. 121–125, 2018.
- [8] S. Majumder, E. Aghayi, M. Noferesti, H. Memarzadeh-Tehran, T. Mondal, Z. Pang, and M. Deen, "Smart homes for elderly healthcare—recent advances and research challenges," *Sensors*, 2017.
- [9] B. Wollschlaeger and K. Kabitzsch, "Navigating the Jungle of Assistance Systems: A Comparison of Standards for Assistance Functionality," in *Proceedings of the 12th BIOSTEC Conference - Volume 5: HEALTHINF*, INSTICC. SciTePress, Feb. 2019, pp. 359–366.
- [10] S. Meyer, R. G. Heinze, M. Neitzel, M. Sudau, and C. Wedemeier, *Technische Assistenzsysteme für ältere Menschen - eine Zukunftsstrategie für die Bau- und Wohnungswirtschaft. Wohnen für ein langes Leben/AAL. Abschlussbericht*, ser. Forschungsinitiative Zukunft Bau. Fraunhofer Irb Stuttgart, 2015.
- [11] M. A. Estudillo-Valderrama, L. M. Roa, J. Reina-Tosina, and I. Román-Martínez, "Ambient assisted living: A methodological approach," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Aug 2010, pp. 2155–2158.
- [12] E. Stav, S. Walderhaug, M. Mikalsen, S. Hanke, and I. Benc, "Development and evaluation of SOA-based AAL services in real-life environments: A case study and lessons learned," *International Journal of Medical Informatics*, vol. 82, no. 11, pp. e293–e293, 2013.
- [13] B. Wollschlaeger, E. Eichenberg, and K. Kabitzsch, "Switching to a holistic perspective on semantic component models in building automation – tapping the full potential of automated design approaches," in *Proceedings of the Sustainable Built Environment D-A-CH Conference 2019*, ser. IOP Conference Series: Earth and Environmental Science (EES), Sep. 2019.

Decreasing Number of Medication Prescriptions After e-Prescriptions Became Mandatory and Their Valid Period Was Extended: A Big Bang Policy Change in Finland in 2017

Vesa JORMANAINEN^{a,1}, Toni RELANDER^a, Vilma JORMANAINEN^b and Marina LINDGREN^c

^a*Finnish Institute for Health and Welfare (THL), Helsinki, Finland*

^b*Äänekoski Town Primary Healthcare Centre, Äänekoski, Finland*

^c*Social Insurance Institution of Finland (Kela), Helsinki, Finland*

Abstract. We measured the first time the number of prescriptions in Finland by using a centralized national shared repository of e-prescriptions and their dispensing data. This was made possible by introducing two policy interventions that were effective since 1 January 2017: mandatory e-prescribing and extension of a prescription's valid period from 12 months to two years. The trend in recording e-prescriptions started decreasing since the 'big bang', but pharmacy dispensations were not affected, as followed up to September 2019. It is possible to monitor the effects of health policy interventions by utilizing national e-health data systems, especially at the occurrence of a natural experiment.

Keywords. Electronic prescribing, large-scale interventions, health policy analysis, national data repositories, eHealth, Kanta services, Finland

1. Introduction

Successful nationwide e-health services, such as electronic prescription service, across different health systems in various countries or regions, can provide valuable information for authorities that have yet to launch these services [1,2].

In Finland, implementation and adoption of the national Kanta services have been carried out in phases step-by-step since May 2010 [3]. Electronic prescriptions (e-prescriptions) are part of the national data system services for healthcare, pharmacies, citizens and social welfare service providers. A fully operational and nationwide shared Prescription Centre has been mandated by law to be implemented in all community pharmacies (2012), in public healthcare (2013), and private healthcare (2015). Finnish community pharmacies started to dispense e-prescriptions in 2012. A web service for physicians and dentists for the issuing of e-prescriptions (Kelain) was launched in

¹ Corresponding Author, Vesa Jormanainen, Social and Health Systems Research, Finnish Institute for Health and Welfare, P.O. Box 30, 00271 Helsinki, Finland; E-mail: vesa.jormanainen@thl.fi.

September 2016. First, in Europe, Finnish citizens could use their e-prescriptions to purchase medication that is dispensed in Estonia (since January 2019) and Croatia (since June 2019).

In the 2007 e-health roadmap [4], Finland chose to ensure the availability of information for patients undergoing treatment, regardless of time and place, in both public and private healthcare. Second, Finland chose to enable the participation of citizens and patients and to ensure that citizens have access to more information, including high-quality health information. These strategic choices are also reflected in the decisions to introduce the national e-prescription system, Prescription Centre, and citizens' access to their health data and e-prescriptions in May 2010.

Finland is a sparsely populated country of 5.5 million inhabitants. In the eastern and northern parts of the country, the population density is especially low, and distances are long. A large number of small municipalities have big responsibilities that include providing both healthcare and social welfare services, which is a unique characteristic of the Finnish healthcare and social welfare system [5,6]. The Finnish health system is highly decentralized and performs rather well, with users expressing comparatively high levels of satisfaction with the quality of services.

Patients are offered access to their health data on a national level only in few countries, such as Australia, Canada, Denmark, Estonia, Finland, France, Iceland, New Zealand, Norway, Scotland, Singapore and Sweden [7]. In Finland, community pharmacy customers are familiar and very satisfied with the national e-health service for viewing their e-prescriptions [8,9]. General practitioners in Finland report that e-prescribing has improved their patient medication management [10,11]. By introducing a nationwide e-prescription and a centralized shared Prescription Centre, Finnish citizens are provided with direct online access to their complete prescription data.

The introduction of a mandatory e-prescribing did not come as a surprise to healthcare actors in Finland. It was discussed lively during Parliament committee hearings in early spring 2014 when Parliament accepted a two-stage introduction of the mandatory e-prescription for the private sector. Since 2017, all public and most private healthcare providers have subscribed and use the Prescription Centre's services. At the same time, the valid period of a prescription was extended to two years.

The administrative nationwide 'big bang' change of 1 January 2017 was essentially constructed on top of a 'blueprint' plan. These large-scale healthcare changes provided us with the first opportunity to measure prescription volume in Finland, and how healthcare professionals (physicians and dentists) learn to use new ways to issue e-prescriptions in their care and cure processes.

2. Material and Method

The Prescription Centre is a national centralized and shared repository of e-prescriptions and their dispensing data. Since January 2017, community pharmacies record paper or telephone prescriptions into the repository (100% coverage).

We extracted prospective monthly indicator data in aggregate form from May 2010 to September 2019. Based on numbers of medication dispensations at community pharmacies in the Finnish Statistics on Medicines, it is estimated that Prescription Centre data was 90% complete in 2015, 95% complete in 2016 and 100% complete after that. Variation in monthly numbers of e-prescriptions and their dispensing events was

smoothed by using a 3-month moving average. In annual calculations of number of persons, one personal identification code was included only once.

3. Results

Totally 162.90 million (M) e-prescriptions and 297.15M dispensations were recorded into the Prescription Centre from May 2010 to December 2018. Altogether 4.34M persons received an e-prescription in 2017 and 4.38M in 2018.

3.1. Medication dispensations in 2015–2019

Numbers of dispensations from e-prescriptions were 49.03M in 2015 (90% coverage; estimate corrected as 54.48M), 55.12M in 2016 (95% coverage; estimate corrected as 58.12M), 61.39M in 2017 and 64.42M in 2018. An estimated corresponding figure in 2019 would be 65.82M. Monthly numbers of dispensations in community pharmacies in Finland, smoothed by using a 3-month moving average, are presented in Figure 1. The trend curve shows an overall increasing time series, and some signs according to which monthly dispensation number vary on average 5.5M dispensations per month.

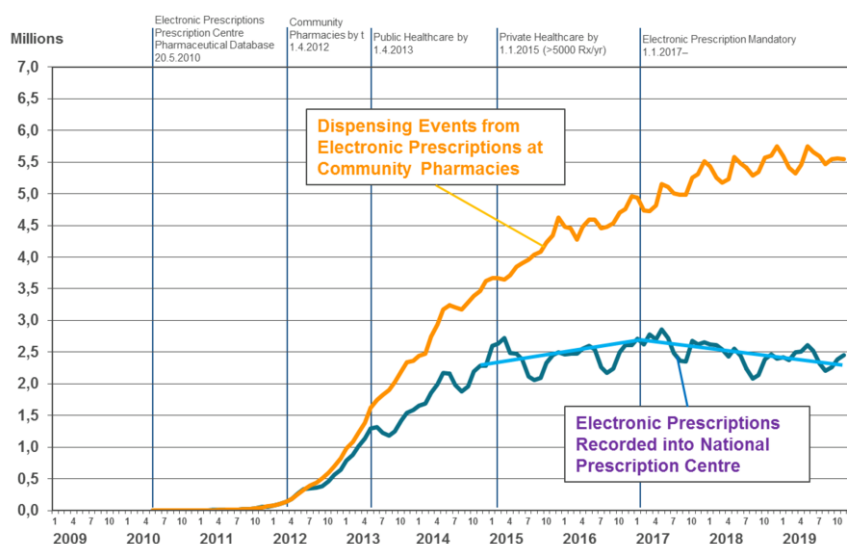


Figure 1. Monthly time series of dispensing events and e-prescriptions from the national Prescription Centre service launch in May 2010 to September 2019 in Finland.

3.2. E-Prescriptions in 2015–2019

Numbers of e-prescriptions were 27.93M in 2015 (if 90% coverage, corrected as 31.03M), 29.61M in 2016 (if 95% coverage, corrected as 32.90M), 31.91M in 2017 and 28.29M in 2018. An estimated corresponding figure in 2019 would be 29.13 M. Annual numbers of recorded e-prescriptions present an increasing trend in 2015–2016, given estimates for 2015–2016 were correct, and a decreasing trend in 2017–2019, given that

the estimate for 2019 was correct. Monthly numbers of recorded e-prescriptions into the national Prescription Centre trend curve shows an overall increasing time series for 2015–2016 and an overall decreasing time series for 2017 onwards (Figure 1). The trend curve also describes an annual decrease of e-prescription recordings in June–July, common summer holiday months in Finland.

3.3. Paper and telephone e-prescriptions in 2017–2018

Numbers of paper and telephone prescriptions recorded at the community pharmacies were 0.404M in 2017 (1.26% of all prescriptions) and 0.384M in 2018 (1.23%). The numbers were highest (on average 37,000 per month) in the first quarter of 2017 but decreased towards end-2018 being on average 31,000 per month during the last quarter in 2018. During 2017–2018, a third of the prescriptions recorded at pharmacies were issued via telephone and two-thirds were paper prescriptions. If one issues a paper or telephone prescription, the reason for action must be recorded. Legal categories include a technical system failure (80% of the reasons) and an emergent need for medication (12.5% on average). The other reasons' proportion decreased from 10% in 2017 to 5% in 2018.

3.4. E-Prescriptions via Kelain web service in 2016–2019

The Kelain web service was launched in September 2016. Using Kelain, the number of e-prescriptions issued was 0.020M in 2016, whereas 0.322M in 2017 and 0.301M in 2018. Monthly numbers rose rapidly to 25,000–30,000 but have since decreased to current 20,000–25,000 level. Comparing numbers of paper prescriptions recorded at pharmacies to those of Kelain web service, it seems that Kelain has taken volumes from paper prescriptions in general. The number of registered Kelain users rose rapidly from 10,000 in early-2017 to current 18,600 in September 2019. The latest number is approximately 50% of the physicians and dentists in Finland.

4. Discussion and Conclusions

We measured the first time number of prescriptions in Finland in 2017 by using national e-health data systems (e.g., Prescription Centre). We were also able to monitor the effects of two simultaneously introduced policy interventions, mandatory use of e-prescribing and extension of a prescription's valid period for up to two years. The trend in recording e-prescriptions decreased since the 'big bang', but pharmacy dispensations were not affected, and this followed up to September 2019.

From a clinician's point of view, e-prescription has brought several welcomed changes. The extended 2-year valid period has reduced amounts of prescription renewals, especially among patients reviewed to be in a stable state, such as steady hypertension. In general, e-prescription is seen as a valuable asset among clinician peers. Physicians can check all the patient's prescriptions through Kanta, and several electronic patient record (EPR) systems are also able to search for and inform the physician on drug interactions otherwise overlooked. The most common complaint heard among peers is that sometimes renewing prescriptions is a labour-intensive process. However, this process varies between EPRs' process designs. Citizens' access to their e-prescriptions

via national My Kanta Pages has also brought possibility for patients to send prescription renewal requests straight to physicians with no accessory contact to their respective healthcare site. With known patients, this reduces the amount of work for a physician. As a downside, this also allows patients to request renewals of prescriptions that are meant for short-term use or prescriptions that require contact to a physician, such as antibiotics, corticosteroids or central nervous system medication. Prescription renewal requests are usually directed to public primary healthcare which may require additional work to be done there.

Large-scale changes (such as described in this paper) allow, at least in theory, for dealing with interrelated elements of a policy. However, such reforms are rare [12]. The outcomes of policy reforms depend on the integration of scale and pace, the intersection of which yields basically four possibilities: rapid enactment of large-scale change ('a big bang'), step-wise enactment of large-scale change ('a blueprint'), rapid enactment of multiple small-scale changes ('a mosaic') or piecemeal enactment of small-scale changes over time (incrementalism). In our case, we first introduced a step-wise enactment of a large-scale change by introducing e-prescription and national Prescription Centre services. In the second stage, based on the results of the implementation, adoption and large-scale use of the Prescription Centre, we were able to change healthcare system and professionals to start using e-prescriptions all over Finland since one date (1 January 2017).

In conclusion, it is possible to introduce large-scale, national, centralized, integrated and shared e-prescription service that is implemented and widely in sustainable use in a country. Full e-prescription coverage (100%) can be used to monitor and evaluate various medication interventions or administrative changes in real-time.

References

- [1] P. Kierkegaard, E-Prescription across Europe, *Health Technol* **3** (2013), 205–219.
- [2] L. Parv, P. Kruus, K. Mötte, P. Ross, An evaluation of e-prescribing at a national level, *Inform Health Soc Care* **41** (2015), 78–95.
- [3] V. Jormanainen, Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study, *Finnish J eHealth eWelfare* **10** (2018), 381–395.
- [4] *eHealth Roadmap – Finland*, Report 2007:15, Ministry of Social Affairs and Health, Helsinki Finland, 2007.
- [5] T. Vehko, S. Ruotsalainen, H. Hyppönen (Eds.), *E-health and e-welfare of Finland: Check point 2018, Report 7/2019*, Finnish Institute for Health and Welfare, Helsinki, Finland, 2019.
- [6] I. Keskimäki, L-K. Tynkkynen, E. Reissell, et al., Finland: Health System review, *Health Systems in Transition* **21** (2019), 1–166.
- [7] C. Nohr, L. Parv, P. Kink, et al., Nationwide citizen access to their health data: analyzing and comparing experiences in Denmark, Estonia and Australia, *BMC Health Serv Res* **17** (2017), 534.
- [8] E. Lämsä, J. Timonen, P. Mäntyselkä, R. Ahonen, Pharmacy customers' experiences with the national online service for viewing electronic prescriptions in Finland, *Int J Med Inform* **97** (2017), 221–228.
- [9] E. Lämsä, J. Timonen, R. Ahonen, Pharmacy customers' experiences with electronic prescriptions: cross-sectional survey on nationwide implementation in Finland, *J Med Internet Res* **20** (2018), e68.
- [10] E. Kivekäs, H. Enlund, E. Borycki, K. Saranto, General practitioners' attitudes towards electronic prescribing and the use of the national prescription centre, *J Eval Clin Pract* **22** (2016), 816–825.
- [11] H. Kauppinen, R. Ahonen, P. Mäntyselkä, J. Timonen, Medication safety and the usability of electronic prescribing as perceived by physicians: a semistructured interview among primary health care physicians in Finland, *J Eval Clin Pract* **23** (2017), 1187–1194.
- [12] C.H. Tuohy, Remaking policy: scale, pace, and political strategy in health care reform, *Studies in comparative political economy and public policy* **54**, University of Toronto Press, Toronto, Canada, 2018.

Exploratory Analysis of HIV Status Knowledge and Associated Factors Using Data from Electronic Medical Records

Natividad BURDISO^{a,1}, Santiago ESTEBAN^a, Karin S. KOPITOWSKI^a and Sergio A. TERRASA^a

^a Family and Community Division, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.

Abstract. Despite recommendations for the routine HIV testing of all sexually active individuals, a significant percentage of HIV-positive adults are unaware of their HIV status. Therefore, a number of strategies have been implemented to expand HIV testing, which in turn makes it necessary to develop tools for identifying patients with unknown HIV status. This study presents the results of an external validation of an electronic phenotyping algorithm for identifying HIV status and its application on a retrospective cohort in order to explore temporal trends of HIV knowledge status and associated factors.

Keywords. HIV, universal HIV testing, electronic phenotyping algorithm, external validation

1. Introduction

The early detection and treatment of human immunodeficiency virus (HIV) infection in its early stages has been shown to have a number of benefits at the both the individual and social level, not only because it decreases the risk of progression to acquired immunodeficiency syndrome (AIDS), but also because it diminishes the risk of sexual and perinatal transmission[1]. Based on these findings, the U.S Preventive Service Task and Force (USPSTF) recommends that clinicians test patients aged 15 to 65 years for HIV infection, as well as all pregnant women regardless of age, whether or not behavioural risk factors have been reported[6]. Nonetheless, a significant percentage of HIV-positive adults are unaware of their HIV status (18% internationally and 30% within Argentina) [3]. Thus, a number of methods have been developed to detect patients with unknown HIV status. Although self-report has been widely used, the validity of this method has been called into question given that patients may not recall whether or not they have previously been tested[5]. Therefore, alternative methods based on information available in clinical records have been designed. Nonetheless, although electronic medical records (EMRs) facilitate the search for HIV-positive patients, who tend to be easily identifiable through laboratory data, the identification of patients with unknown HIV status is more complex since there is currently no reliable method for assigning a patient to that category. Accordingly, unknown HIV status should be inferred from the

¹Corresponding Author, E-mail: natividad.burdisso@hospitalitaliano.org.ar

absence of information identifying a patient as HIV-positive or HIV-negative. Based on this assumption, Felsen et al. [4] developed a rule-based algorithm for electronic phenotyping based on electronic medical records in order to classify patients by their HIV status. The algorithm's accuracy for identifying patients with unknown HIV status was assessed by comparing the results to a manual review. Its sensitivity and specificity were 99.4% (95% CI: 96.5-100%) and 95.2% (95% CI: 83.8-99.4%) respectively. Despite the fact that the development of EMR-based algorithms has become popular in recent years, their external validation – that is, their application to populations other than those they were initially developed for – is far less common [7]. Thus, we decided to externally validate the algorithm developed by Felsen et al. using data obtained from the EMR at Hospital Italiano de Buenos Aires (HIBA) and apply it to a cohort of patients to explore the temporal trends of HIV status knowledge change and associated factors. The two main aims of our study are 1) To locally validate the EMR-based phenotyping algorithm for identifying HIV status developed by Felsen et al. and 2) To describe the trends over time in the percentage of individuals who know their HIV status among patients enrolled in the health insurance plan at HIBA and explore the relationship between HIV knowledge status change, sex, age and year of enrollment.

2. Methods

For the validation of the phenotyping algorithm, two types of variables were included: 1) standardized codes linked to HIV/AIDS (used at inpatient, outpatient, and emergency department facilities); and 2) laboratory data such as anti-HIV antibodies, western blot, viral load, and CD4 cell count. No free-text data was used. The algorithm is shown in figure 1.

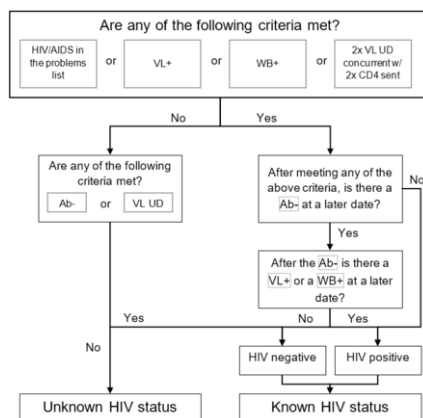


Figure 1. HIV status classification algorithm. VL: Viral load; WB: Western Blot; UD: undetectable; Ab: ELISA antibody test.

Manual chart review was defined as the reference standard. Two family physicians analyzed all of the data available in the EMR of each patient. The criteria for classifying patients according to their HIV status were as follows: HIV-positive: presence of a positive WB, detectable VL, documented HIV infection by a provider, prescription of antiretroviral therapy, or mention of infection in narrative text notes; HIV-negative: does not meet criteria for HIV-positive, and the presence of a negative western blot or ELISA

test for anti-HIV antibodies; Unknown HIV status: does not meet criteria for HIV-positive or HIV-negative. Assuming that our algorithm would have a sensibility of 99% and a specificity of 95.2% – as in Felsen et al. (2014) – a sample size of 300 observations was estimated in order to obtain a margin of error of $\pm 0.75\%$ on a 95% confidence interval in our test set.

For the assessment of HIV status temporal trends, we applied the algorithm to all the patients enrolled in the insurance plan at the HIBA for at least 12 months aged 15 to 65 between 1/1/2008 and 31/12/2015, every year at the last day of the year. We measured the percentage of HIV status knowledge both per year and cumulative. We also assessed the percentage of patients that changed their status over time. Lastly, using a three year follow up sub-cohort of yearly newly enrolled patients (assuming their status before enrolment was ‘unknown’), we assessed the time until status change using a cox proportional hazards model which allowed us to explore its association with sex, age and cohort entry year. All statistical analyses were computed using R.

3. Results

When applied to our test set, the algorithm showed similar metrics to those reports by the authors [sensitivity 97.25% (92.16 - 99.42); specificity 100% (98.08 - 100), positive predictive value 100% (97 - 100) and negative predictive value 98.45% (96 - 100)]. regarding HIV knowledge status trends, as it is shown in table 2, ‘known’ HIV status (positive + negative results) percentage has been increasing since 2008, mostly due to an increasing percentage of negative tests, at a mean increase of 0.6% per year. HIV knowledge status change (from ‘unknown’ to ‘known’) appears to be in an upward trend since 2013, which coincides with the initial USPSTF recommendation (figure 2). This trend seems to be led by young adults (20 – 40 years old). Upon visual inspection of the trends between sexes, a higher percentage of women have changed their status to ‘known’, although men show a steeper curve since 2013.

Table 2. Patients’ description and HIV knowledge status.

Year	n	Sex - Female (%)	Age (median [IQR])	Known status – Negative (%)	Known status – Positive (%)	Unknown status (%)	At least one HIV test since 2008 (%)	Newly enrolled (%)
2008	89481	49340 (55.1)	40.00 [29.00, 54.00]	6164 (6.9)	234 (0.3)	83083 (92.8)	19082 (21.3)	11352 (12.7)
		49977	40.00	6393	240	83988	20892	9703
2009	90621	50656 (55.1)	40.00 [29.00, 54.00]	6633 (7.1)	115 (0.3)	85064 (92.7)	22732 (23.1)	9346 (10.7)
		50656	40.00	6633	115	85064	22732	9346
2010	91812	48793 (55.2)	42.00 [30.00, 54.00]	6912 (7.2)	126 (0.1)	79378 (92.7)	23601 (24.8)	9655 (10.2)
		48793	42.00	6912	126	79378	23601	9655
2011	86416	50167 (56.5)	41.00 [31.00, 55.00]	7228 (8.0)	160 (0.1)	81701 (91.9)	25591 (27.3)	9966 (11.2)
		50167	41.00	7228	160	81701	25591	9966
2012	89089	50291 (56.3)	42.00 [31.00, 55.00]	7761 (8.1)	161 (0.2)	81618 (91.7)	27498 (28.7)	8028 (11.2)
		50291	42.00	7761	161	81618	27498	8028
2013	89540	49773 (56.2)	42.00 [31.00, 55.00]	8726 (8.7)	150 (0.2)	80204 (91.2)	29591 (30.7)	7139 (9.0)
		49773	42.00	8726	150	80204	29591	7139
2014	89080	49527 (55.9)	42.00 [31.00, 55.00]	9981 (9.8)	155 (0.2)	78936 (90.0)	32321 (33.2)	7257 (8.0)
		49527	42.00	9981	155	78936	32321	7257
2015	89072	(55.6)	[31.00, 54.00]	(11.2)	(0.2)	(88.6)	(36.3)	(8.1)

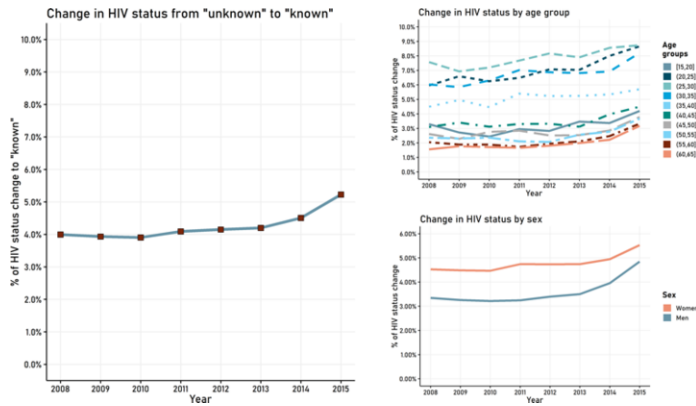


Figure 2. HIV knowledge status change.

Figure 3 and 4 show a clear and independent association between sex, age, cohort entry year and time until status change. Cohort entry year shows to be strongly associated with a higher proportion of patients who changed their status from ‘unknown’ to ‘known’ after the three years of follow up. This association seems stronger for every year compared with 2008, as is shown by the increasing hazard ratios.

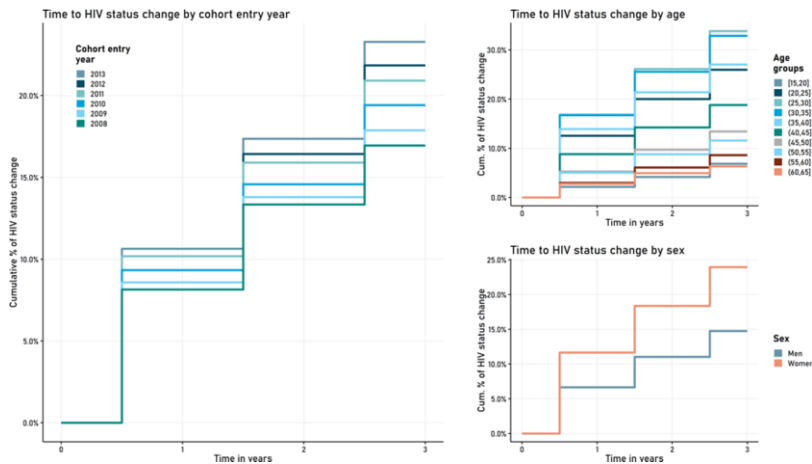


Figure 3. Unadjusted survival curves for time until HIV knowledge status change in the 3-year follow up sub-cohort of newly enrolled patients.

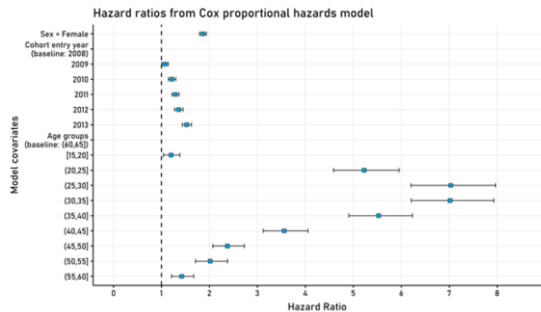


Figure 4. Hazard ratios from Cox proportional hazards model exploring the association between time until HIV knowledge status change, age, sex and cohort entry year in the 3-year follow up sub-cohort of newly enrolled patients.

4. Discussion

When applied to our population, the performance characteristics of the original algorithm were similar to those initially reported by the authors. EMRs combined with validated algorithms provide a useful source of information for monitoring temporal trends.

In our results, patients between 20 and 40 years of age, showed to be consistently more likely to have at least one HIV test and thus change their HIV knowledge status from 'unknown' to 'known'. This, combined with the overall upward trend of status change over the years that we found, is promising for the future given the clear benefits of early diagnosis that have been demonstrated over the last decade[1]. Nevertheless, the lower proportion of older adults that have had access to an HIV test in our study population is of concern given the higher comorbidity and mortality associated to older adults with delayed HIV diagnosis[2].

5. Conclusion

We were able to validate the algorithm developed by Felsen et al. in our population, which showed to have similar performance metrics as the ones reported the authors. When applied to our cohort, we found that HIV status knowledge change has been increasing, particularly since 2013 and that it seems to be independently associated with sex, age and cohort entry year.

References

- [1] M.S. Cohen, Y.Q. Chen, M. McCauley, T. Gamble, M.C. Hosseinipour, et al., Prevention of HIV-1 infection with early antiretroviral therapy, *N Engl J Med* **365** (2011), 493-505.
- [2] D.H. Davis, R. Smith, A. Brown, B. Rice, Z. Yin, and V. Delpech, Early diagnosis and treatment of HIV infection: magnitude of benefit on short-term mortality is greatest in older adults, *Age Ageing* **42** (2013), 520-526.
- [3] Boletín sobre el VIH-SIDA e ITS en la Argentina, 112, 2013.
- [4] U.R. Felsen, E.Y. Bellin, C.O. Cunningham, and B.S. Zingman, Development of an electronic medical record-based algorithm to identify patients with unknown HIV status, *AIDS Care* **26** (2014), 1318-1325.
- [5] S.M. Jenness, C.S. Murrill, K.L. Liu, T. Wendel, E. Begier, and H. Hagan, Missed opportunities for HIV testing among high-risk heterosexuals, *Sex Transm Dis* **36** (2009), 704-710.
- [6] D.K. Owens, K.W. Davidson, A.H. Krist, M.J. Barry, M. Cabana, et al., Screening for HIV Infection: US Preventive Services Task Force Recommendation Statement, *Jama* **321** (2019), 2326-2336.
- [7] E.W. Steyerberg, K.G. Moons, D.A. van der Windt, J.A. Hayden, P. Perel, et al., Prognosis Research Strategy (PROGRESS) 3: prognostic model research, *PLoS Med* **10** (2013), e1001381.

Global Public Health Surveillance Using Media Reports: Redesigning GPHIN

Dave CARTER ^a, Marta STOJANOVIC ^a, Philip HACHEY ^a, Kevin FOURNIER ^a,
Simon RODIER ^a, Yunli WANG ^a, and Berry DE BRUIJN ^a

^a *Digital Technologies Research Centre, National Research Council Canada, Ottawa*

Abstract. Global public health surveillance relies on reporting structures and transmission of trustworthy health reports. But in practice, these processes may not always be fast enough, or are hindered by procedural, technical, or political barriers. GPHIN, the Global Public Health Intelligence Network, was designed in the late 1990s to scour mainstream news for health events, as that travels faster and more freely. This paper outlines the next generation of GPHIN, which went live in 2017, and reports on design decisions underpinning its new functions and innovations.

Keywords. Public Health, Syndromic Surveillance, Natural Language Processing

1. Introduction

Global public health surveillance relies on reporting structures and transmission of trustworthy health reports. But in practice, these processes may not always be fast enough, or are hindered by procedural, technical, or political barriers. At the same time, the news travels much faster and more freely, certainly in this day and age. Recognizing that, Public Health Agency of Canada has been operating the Global Public Health Intelligence Network (GPHIN) since 1997, designed around capturing mainstream media stories from around the world, analyzing them, and then disseminating the insights globally, through WHO, for rapid responses such as initiating treatment and vaccination efforts, and slowing spread through travel advisories and port-of-entry screening.

This paper describes the efforts and results from the GPHIN rejuvenation project.¹ It provides an overview of the (added) functionalities, design decisions, and high-level implementation details. It also aims to inspire readers that are looking to extend these ideas to other types of real-time, broad scope surveillance efforts in life sciences.

2. Background

For a detailed overview of disease surveillance principles and systems, we refer to review papers on this topic. Abat et al. in their 2016 review paper [1] put multiple syndromic surveillance efforts in perspective, citing multiple data sources from the environment (including weather and animal data), population behaviours (including drug sales and

¹ GPHIN renewal was a team effort. The authors wish to acknowledge the contributions of research, development, operational, and support staff at National Research Council and Public Health Agency of Canada.

inter- net use), health care (clinics and hospitals), and demographics. We note that mainstream media was not considered in this review. Barboza et al. [2] analyzed several earlier bio- surveillance systems and concluded that they are indeed generally effective at publishing information about infectious diseases using news media compared to a goldstandard.

In addition to human disease surveillance, there are monitoring systems for livestock health that can inspire us. Lower privacy concerns surrounding animals have contributed to quicker advancements of such systems. Additionally, problems with livestock health can lead to human health concerns downstream, as animals can function as human disease vectors, and livestock epidemics can affect the human food supply chain [9]. EpidVis [7] is one such initiative that combines news collection with compelling visualizations for animal epidemiology.

GeoSentinel is a syndromic surveillance network that collects and disseminates reports on sicknesses in travellers. The information can inform pre-travel preparation, as well as guide post-travel differential diagnosis when a recent traveller has fallen ill [8].

The first version of GPHIN was conceived in the late 1990s, and soon proved its worth by enabling early detection of the 2003 SARS outbreak [6]. While the premise of the system remained relevant, advancements in software technology and data processing methods warranted a ground-up rejuvenation effort, which this paper reports on.

3. Methods and Results

The base architecture of GPHIN involves data intake, data processing, information storage, and interactive user tools [3]. In many cases, third-party or off-the-shelf components are used. Data originates from multiple sources, the majority of articles arriving through a large media aggregation service, others through curated RSS feeds – news sites, public health organizations, product recalls, and some high-relevance epidemiological (micro)blogs. Processed and tagged data are then stored in a relational database and also in an off-the-shelf full-text search tool, for analysts to search, retrieve, and digest.

A number of innovations were implemented. We highlight several of these in the following paragraphs, starting with its multilingual capabilities, including translation.

Multiple languages: GPHIN is a global effort, and as it uses local-level mainstream media signals, it must be able to process multiple languages. GPHIN processes data in ten languages: English, French, Spanish, Portuguese, Russian, Arabic, Farsi, Simplified and Traditional Chinese, and Indonesian – covering logograph, abjad and alphabet scripts and covering both left-to-right and right-to-left languages. Text analytics takes place partly in the language of origin (keyphrase extraction from restricted phrase lists), partly based on English machine-translations of the source document. Pragmatically, the Microsoft Azure Translate API was selected to detect the language of source material and translate it to English. These translations are somewhat noisy, and are never used as full substitutes to the source text, but even lossy translation can inform human analysts as well as various machine analytics components such as classifiers. Occasionally, downstream problems such as ambiguity can be averted by using modern context-sensitive machine translation algorithms: for example, the Chinese character can mean *twenty-first*, the family name *Ma*, or *horse*, and only one meaning is relevant when tracking equine diseases.

Relevance scoring is a function that discriminates first-line reports from background stories, retrospectives, opinions, policy documents, or any story that was accidentally caught by the aggregator service (e.g., “Bieber fever”). The technique is implemented as a binary classifier but with a confidence estimate that allows for ranking and flexible threshold setting. The classifier [5] is built around a machine learning classifier (SVM) using a set of approximately 1400 documents represented as unrestricted word and character n-grams. Articles with a high relevance score are published immediately, while low-scoring articles are automatically suppressed; only the remaining medium-relevance articles are triaged by analysts.

Information extraction: Medical terms in text are detected and mapped to UMLS by MetaMap Lite. The tool is restricted to a set of semantic types that is generally relevant to the public health community – largely syndromes and disease vectors. Subsequently, a hand-curated blacklist of common false positive terms is applied. The resulting annotations inform downstream components such as clinical reasoning, and links to UMLS definitions and ontology elements are provided next to articles. General named entities, such as person and organization names, are tagged using Stanford CoreNLP’s named entity recognizer. Heuristics were created to infer people’s job titles when mentioned, as well as to consolidate organizations’ full names and acronyms.

Geographic entity resolution: One part of situational awareness is knowing precisely where things are happening. We use Stanford CoreNLP’s named entity recognition plus heuristics to resolve mentions of geographic entities (countries, states/provinces, cities/towns, some lakes, some regions) to latitude-longitude coordinates and to canonical forms in the Geonames ontology so that they can be counted and mapped, collecting synonymous mentions of places and inferring parent-child relationships. When large outlets report on Tokyo or New York, local news sites may specify the sub-city (Yokohama or Shinjuku; Queens or Tribeca). A search in GPHIN for Tokyo will include all sub-cities, which makes it easier to synthesize very local media from a global perspective. This function also reconciles synonyms and language differences, coalescing multiple spellings/forms into a single known entity. Non-unique place names (e.g. London, UK versus London, Ontario, Canada) are resolved through heuristics that take into account where an article was published. Geo-tagging allows users to quickly drill into details about the geography of a document, checking population sizes, median age, poverty rates, and human development index measurements, for example.

Deduplication: Near-duplicate reports on the same incident are unavoidable when many media outlets republish newswire articles verbatim. To avoid duplicate stories hijacking trend detection algorithms, a duplicate detection algorithm was devised. The technique is built on two main principles: text edit distance, where a certain percentage of the word triplets in the article must be the same; and shingling, an observation that if all possible three- or four-word phrases are randomly assigned a number, sorting and sampling the set of such unique numbers for the phrases in two documents produces a substantially similar sample for documents that are similar but not necessarily identical. This solves the challenges like editors lightly rewording newswire articles for space constraints and adding/removing extra paragraphs of context at the end of an article. Duplicates are not removed but clustered / flagged as such, and the initial version among candidate duplicates is used as the exemplar. Minor but critical differences, such as an updated death count, are extracted using StructPred and tracked separately.

Narrative discovery and change point detection: On a daily basis, it looks for statistically unusual pairs of non-generic ontology keywords plus locations, and publishes these in a list of new “narratives”. As new articles come in to the system that

match each such pair of keyword plus location, it aggregates the content and tracks the narrative. While GPHIN may ingest as many as 10 000 new articles in a given day, the number of new narratives tends to be between 20 to 50 daily, which is small enough for an analyst to digest. For narratives that span at least several days, GPHIN applies topic modelling technology [10] to estimate when the language of the narrative is changing – when the language of an infectious disease pandemic is changing from uncertain to certain, or from investigation to response to rehabilitation, and so on. While GPHIN was designed around data streams that carry more noise than permissible for highly sensitive aberration detection algorithms, this combination of narrative discovery plus change point detection accomplishes a similar goal of trend detection and tracking.

User interface: The dynamic and faceted nature of the task is supported by a range of user interface modes. Figure 1 illustrates some of these. The basic interaction model in GPHIN is a set of filtering criteria coordinated with a colour-coded map, a configurable bar graph (showing activity over time or top keywords/locations/categories), and a list of documents. If a user has a hypothesis to test, a user might compose a search for a tagged syndrome in a geographic area, and then drill down by clicking on red hot spots on the interactive map or drill down into increasingly specific keywords or combinations thereof (Figure 1, left). If the event is a serious outbreak or a mass gathering, there are integrated report generation tools to support daily surveillance workflows outside the tool – users can tag and collate documents into topical reports, and GPHIN analysts can highlight portions of text within documents. Analysts can disseminate their findings via e-mail and RSS, while recipients can specify their own alert filters.

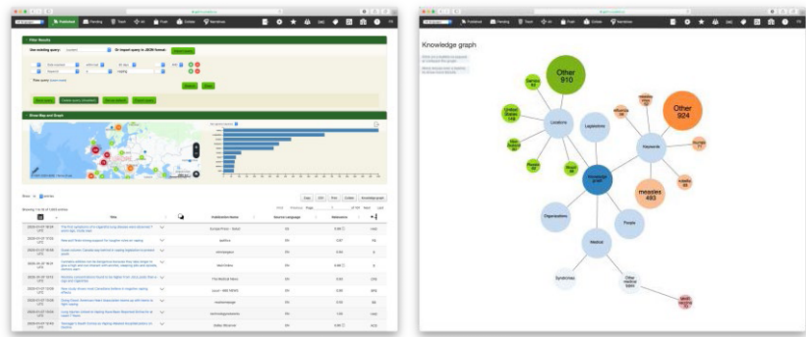


Figure 1. Coordinated search, map, and graph tool (left); knowledge graph for a measles outbreak (right)

GPHIN also supports workflows where the user has no specific hypothesis. The recently discovered narratives are listed with recent activity time series graphs, extractive summaries [11], and any detected inflection points. A further tool that GPHIN provides to try to illustrate surprising correlations is a basic knowledge graph (Figure 1, right), placing top keywords, locations, people and medical entities in bubbles. Location bubbles are linked when they are geographically adjacent. GPHIN is used primarily for monitoring current events in recent news, but also supports retrospective research, whether for psychoactive substances [12] or for analysing how media reporting changed the behaviour of the general public in the midst of past pandemics [4].

4. Discussion

GPHIN currently has more than one thousand registered user accounts in 56 countries, representing public health, governments at various levels, military, research, education, diplomatic, and travel organizations, from global to municipal. Aside from day-to-day syndromic surveillance activities, GPHIN was used for specific mass gatherings and summits (Olympics, G7), as these tend to come with exceptional volumes and directions of human (and therefore disease) travel. Surrounding the 2016 summer Olympic Games in Rio de Janeiro, WHO epidemiologists were not just concerned about Zika, but also that an ongoing MERS outbreak in the Middle East might migrate to Rio.

As domain knowledge was kept separate from the software code, we have since expanded the platform to support food safety surveillance (food, animal, and plant hazards, for example) and natural disasters in the context of public safety.

5. Conclusion

We presented an overview of added functionality to a global syndromic surveillance tool that leverages local mainstream media articles, in a range of different languages, to provide a rapid and global view to public health officials.

References

- [1] C. Abat et al., Traditional and syndromic surveillance of infectious diseases and pathogens. *Int. J. Infect. Dis.*, 48:22–28, Jul 2016.
- [2] P. Barboza et al., Factors influencing performance of internet-based biosurveillance systems used in epidemic intelligence for early detection of infectious diseases outbreaks. *PLoS ONE*, 9(3):e90536, 2014.
- [3] D. Carter, Stojanovic M, and de Bruijn B. Revitalizing the Global Public Health Intelligence Network (GPHIN). *Online J. Public Health Inform.*, 10(1), 2018.
- [4] S. Collinson, K. Khan, and J.M. Heffernan, The Effects of Media Reports on Disease Spread and Important Public Health Measurements. *PLoS ONE*, 10:e0141423, 2015.
- [5] B. de Bruijn, M.S. Marshall, and A. Dekker, Fast and simple text classification gets things done. In Scott PJ et al., editors, *Informatics for Health 2017: Advancing both science and practice*. 2017.
- [6] M. Dion, P. AbdelMalik, and A. Mawudeku, Big Data and the Global Public Health Intelligence Network (GPHIN). *Can. Commun. Dis. Rep.*, 41(9):209–214, Sep 2015.
- [7] S. Fadloun et al., Integration of text-and web-mining results in epidvis. In *International Conference on Applications of Natural Language to Information Systems*, pages 437–440. Springer, 2018.
- [8] K. Leder et al., GeoSentinel surveillance of illness in returned travelers, 2007–2011. *Ann. Intern. Med.*, 158(6):456–468, Mar 2013.
- [9] J.P.S. Neo and B.H. Tan, The use of animals as a surveillance tool for monitoring environmental health hazards, human health hazards and bioterrorism. *Vet. Microbiol.*, 203:40–48, May 2017.
- [10] Y. Wang, and C. Goutte, Real-time change point detection using on-line topic models. In *Proceedings of COLING 2018*, pages 2505–2515, 2018.
- [11] Y. Wang, Y. Jin, X. Zhu, and C. Goutte, Extracting discriminative keyphrases with learned semantic hierarchies. In *Proceedings of COLING 2016*, pages 932–942, 2016.
- [12] M. Young, C. Dubeau, and O. Corazza, Detecting a signal in the noise: monitoring the global spread of novel psychoactive substances using media and other open-source information. *Hum. Psychopharm. Clin.*, 30(4):319–326, 2015.

ICT Toolkit for Integrated Prevention, Testing and Linkage to Care Across HIV, Hepatitis, STIs and Tuberculosis in Europe

Christine KAKALOU^{a,1}, Lella COSMARO^b, Vlasios K. DIMITRIADIS^a,
Anne RAAHAUGE^c, Dorthe RABEN^c and Vassilis KOUTKIAS^a

^a *Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thessaloniki, Greece*

^b *"Fondazione LILA Milano" - Italian League for Fighting AIDS, Italy*

^c *CHIP, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark*

Abstract. Online digital tools are considered an innovative method to promote HIV, hepatitis and STIs prevention, testing and treatment services, overcoming individual and social barriers, especially for younger people and other, possibly hard-to-reach, target population groups. In this paper, we introduce INTEGRATE RiskRadar, a web and mobile application developed in the scope of the EU-supported INTEGRATE Joint Action (JA), that aims to enhance the integration of combination prevention, testing and linkage to care for HIV, hepatitis, STIs and tuberculosis by providing integrated information and digital tools regarding all four diseases to population groups at increased risk, aiming to eliminate the individual and social barriers to effective adoption of prevention practices, testing and linkage to care, and thus reduce the incidence and burden of these diseases in the European Region.

Keywords. Human immunodeficiency viruses (HIV), hepatitis, sexually transmitted infections (STIs), tuberculosis (TB), integrated approach, prevention, combination prevention, testing, linkage to care, eHealth, mHealth.

1. Introduction

Despite the fact that the global HIV response has resulted in impressive achievements, HIV epidemics continue to have a profound impact on the health and lives of millions of people worldwide [1]. At present, the coverage of the HIV services is still inadequate, and the rate of their expansion has to be accelerated to meet the 2020 target as well as the 2030 sustainable development goals (SDGs). Equally important, one of the key challenges is the burden of HIV coinfections, namely sexually transmitted infections (STIs), viral hepatitis and tuberculosis (TB), and the impact they have on HIV prevention, treatment and the related increasing morbidity and mortality [1]. This calls for the implementation of several prevention strategies that, if combined, may have additive and synergistic effects on reducing new HIV transmission incidences. Combination prevention is the strategic and coordinated use of biomedical, behavioural and social/structural activities operating on various levels (individual, community, societal) to reinforce the sustained impact of HIV prevention. The World Health Organization

¹ Corresponding Author: Christine Kakalou, Institute of Applied Biosciences, Centre for Research and Technology, Thessaloniki, Greece; E-mail: ckakalou@certh.gr.

(WHO) emphasises that the existing momentum of the HIV response should be bolstered by new evidence- and rights-based combination prevention tools that are tailored to the needs of specific people and communities, especially those of vulnerable population groups that are disproportionately affected by HIV [1].

In this context, digital tools are considered an innovative method to promote HIV, hepatitis and STIs prevention, testing and treatment and overcome individual and social barriers, especially for young people and hard-to-reach target population groups [2], given the rapid uptake and ubiquity of smartphones and Internet access in Europe. Although notable progress has been achieved recently in the fields of eHealth and mHealth, the main focus has been on the management of non-communicable diseases (NCDs), rather than infectious ones. While a growing pool of HIV/hepatitis/STIs-related apps and online tools exists [3], they mostly target medication adherence and specific groups [4]. Several gaps remain in the implementation agenda, e.g. the inclusion of more key populations and other points along the care continuum besides treatment [4] [5].

In this paper we introduce INTEGRATE RiskRadar, a web-based and mobile application developed in the scope of the INTEGRATE joint action (JA), that aims to enhance the effectiveness of combination prevention for HIV, hepatitis, STIs and TB. By providing integrated information and digital tools regarding all four diseases to the public regardless of gender, age, sexual orientation or insurance status, we aim to eliminate the individual and social barriers to effective adoption of prevention practices, testing and linkage to care, and to thus reduce the incidence and burden of these diseases in EU member states.

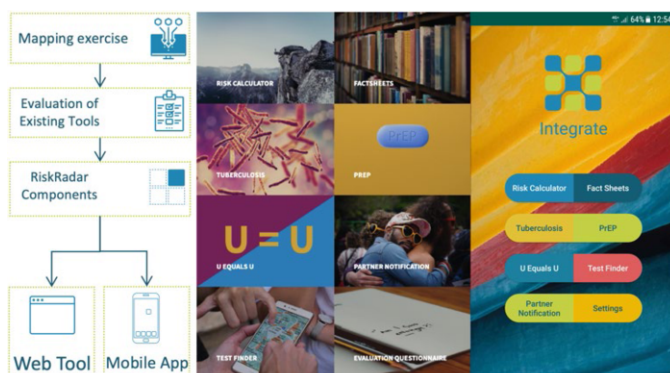


Figure 1. Overall *RiskRadar* development process and landing page for Web tool and mobile app.

2. Methodology

Our design and development approach focused on the assessment of available relevant digital tools to identify promising and innovative features and their consequent adaptation to address all four disease areas. The design of this new integrated ICT (information and communication technology) toolkit accommodates the distinctive needs of various vulnerable groups, i.e., *Young adults* and *men who have sex with men (MSM)* who adopt new partner-seeking and sexual behaviours [6], might engage in illicit drug use [7], report an enhanced need for privacy and anonymity, especially in some countries and demand an attractive and modern user interface (UI) [8]. For *persons who inject drugs (PWID)* risk reduction advice is needed, while *migrants* require a toolkit

with mobility and ubiquity to help them acquire an understanding of additional risk factors. Finally, *prisoners and homeless persons* would benefit from an easily accessible, fast and simple web interface available at clinics and NGO sites, focusing on risk reduction. The RiskRadar development methodology involved three phases (**Figure 1**).

2.1. Mapping Exercise

First, a mapping exercise was conducted, aiming to identify existing ICT resources addressing combination prevention strategies worldwide with additional focus on tools implemented on a national level in the European region. A "mapping matrix" was introduced to align the data collection process, where a specific structure was proposed, featuring fields describing the tool, the target audience, the disease(s) addressed and its availability. INTEGRATE partners were invited to contribute to this phase by submitting any relevant tool they were aware of. The outcomes of the review were harmonised, resulting in the complete mapping matrix of ICT tools, consisting of 115 candidates.

2.2. Evaluation of Existing Tools

To evaluate the collected tools, we determined the appropriate criteria by analysing the matrix and excluding any non-ICT tools such as online guides, manuals or training modules. These were flagged for further consideration during the JA's capacity building activities. The final list of tools for consideration was reduced to 53 and divided into two groups, according to their respective uses in the care continuum, i.e., **Group A-Prevention**, containing 29 tools related to prevention, harm reduction and PrEP (pre-exposure prophylaxis); and **Group B-Testing**, containing 24 tools related to testing, partner notification and linkage to care. A two-fold, more thorough review of these tools followed based on the following criteria: (a) perceived relevance and adaptability to the JA purposes, and (b) adaptability, reproducibility, and data relevance from a technical standpoint, since the vast majority of the tools addressed a single disease.

The review process identified seven tools that scored high in all criteria, then the results were discussed with respect to their strengths, weaknesses and shortcomings with the JA steering committee members. An initial proposal for the RiskRadar structure was formulated according to the toolkit's intended use, target audiences, and overall scope.

2.3. RiskRadar Components

The proposed toolkit features modular components that could be enabled according to the capacity and intended target audience. The following components were defined: (a) a **Factsheets** component offering basic information regarding the different diseases; (b) a **Risk Calculator** (RC) tool that allows users to assess if they have been exposed to the diseases based on underlying risk assessment algorithms, tailored to each key group's needs; (c) a **PrEP** component to inform the user of this novel prevention tool; (d) a **Test Finder** featuring filtering capabilities per disease, type of point-of-care, and country; (e) a **Reminders** component to help users stay on track with their respective testing schedule; (f) a **Partner Notification** (PN) tool based on the combination of existing services, and (g) a **TB** component offering information and highlighting the differences with the other three disease areas in terms of transmission, prevention, and care. RiskRadar will be piloted in Croatia, Italy and Lithuania; thus appropriate translations of its components have been prepared to accommodate language barriers.

Risk Calculator

Are you vaccinated for hepatitis B?

Yes No I don't know

← Risk Calculator

Integrate

Why are you using this risk calculator?

☐ I am HIV+

☐ My partner is HIV+

☐ Both me and my partner are HIV+

☐ I had condomless sex with someone

☐ I inject drugs

☒ I'm just curious

CONTINUE

Result:

Condoms when used always and correctly are very effective at preventing HIV and reducing a risk of STI. Take the Condom Test to help you choose the right condom for you. Other sexual acts such as fisting and sharing sex toys without cleaning them afterwards can also expose you to hepatitis and STIs. It is recommended to get tested at least twice per year or more often if you have sex often and with different partners.

You might have been exposed to HIV and hepatitis through shared injecting equipment; you should get tested. Remember that injecting or using drugs exposes your health to other risks and harms. If you would like to stop using drugs but cannot make it on your own, it is important that you seek help. You can also seek advice and learn about harm reduction strategies, in order to reduce as much as possible the risks associated to drug use.

It is recommended that all gay and bisexual men get vaccinated for Hepatitis A & B; consider asking your medical doctor for a Hepatitis A & B vaccination (You will need a test before getting vaccinated).

Screening for hepatitis C is recommended if you come from Central Asia; Asia; Sub-Saharan Africa; Middle-East; Central and South America where the virus is highly prevalent.

Tuberculosis is highly prevalent in Eastern Europe, Sub-Saharan Africa or Asia – so if you originate from these regions it is recommended to consult a medical doctor for TB screening.

Got it! Retake quiz Take the condom test

Figure 2. (a) Question of the RC component in the web app (upper left part) and question from the RC component in the mobile app (lower left part); (b) the result of the RC (the right part).

2.4. Design and Development Considerations

The main challenge concerned the RC component, since (to the authors' knowledge) a comprehensive tool for calculating the inextricable nature of the combined risks for all four disease areas tailored to the high-risk key populations consisting RiskRadar's end users is missing. A consolidation of all available guidelines by WHO and ECDC reinforced the algorithms derived by the previous analysis of the partial risk assessment tools for each disease. The RC decision tree resulted from an extensive iterative process of obtaining input from all partners, while its final release was approved for official translation in the piloting countries' languages. As regards the information components, contacts with central information sources were established, such as NAM for the Test Finder, "PrEP in Europe" for the PrEP component and ECDC for the factsheets.

Overall, the design approach relied on (a) *effective communication strategies* outlined by WHO and ECDC, and (b) *informed UI design* based on the literature findings of relevant websites and mobile apps [9]. Emphasis was placed on offering adequate information to the user to make informed decisions and possibly adjust risky behaviours while keeping users with low health/eHealth literacy engaged with the app.

3. Results

RiskRadar is available as a web application and as native Android and iOS mobile apps, integrating the various components presented in section 2.3. Overall, our development approach was based on the following pillars: (a) *easy to update content* and (b) *adaptable content to fit diverse target groups*, via a modular design for the RC answers which can be independently modified according to the target audience (Figure 2); (c) *easy to use, intuitive and interactive UI*; (d) *use of acceptable, non-judgmental, non-threatening and non-fear-inducing language and imagery*, through an iterative process of consulting with various public health institutions and relevant NGOs; (e) *patient empowerment* by providing direct PN services as well as a self-evaluation risk calculator tool, and (f) *ensure that the user feels safe and also that their trust, privacy, and confidentiality are clearly perceived while using the app*. The latter is achieved by establishing privacy and cookie policies that explicitly state the anonymisation and encryption mechanisms

employed, featuring them prominently in the application; namely the use of anonymous, unique random codes to access the PN component, the encryption and enhanced security during data transfer and the absence of any identifiable information in the RC component.

Each pilot country targeted different populations. Currently, the pilot sites are testing RiskRadar before its release to the public. The outcomes of the pilots will be used to evaluate RiskRadar's effectiveness and usability. A short questionnaire was added to support pilot exercise, given that it is difficult to evaluate the effectiveness of ICT tools' in the prevention of the four diseases. This is because measures such as the number of web accesses or downloads do not give a precise measure of the usefulness of ICT tools. Nonetheless, RiskRadar has the potential to be an appealing source of information for the prevention and testing of these diseases. Especially for people who encounter barriers to care, RiskRadar could facilitate entry to and retention in the healthcare system.

4. Conclusion

In this paper, we presented the overall approach regarding the design and development of an integrated toolkit accessible over the web or a dedicated mobile application, in line with the current public health authorities' guidelines. The development of RiskRadar exploits the expertise and various insights from INTEGRATE's inter-disciplinary consortium, thus enabling the synergy of biomedical and behavioural prevention and offers a new level of privacy, confidentiality and security as well as the opportunity to satisfy the diverse needs of various target groups at the same time (scalability).

5. Acknowledgement

This work has received funding from the European Union for the INTEGRATE Joint Action within the framework of the Third Health Programme (grant number 761319).

References

- [1] *Global health sector strategies on HIV, viral hepatitis and sexually transmitted infections, for the period 2016–2021*, World Health Organization, 2016.
- [2] *Novel approaches to testing for sexually transmitted infections, including HIV and hepatitis B and C in Europe*, European Centre for Disease Prevention and Control, 2012.
- [3] J. Gibbs, et al., "Can you recommend any good STI apps?" A review of content, accuracy and comprehensiveness of current mobile medical applications for STIs and related genital infections., *Sex. Transm. Infect.* **93** (2017), 234–235.
- [4] J.M. Simoni, et al., Opportunities and Challenges of Digital Technology for HIV Treatment and Prevention, *Curr. HIV/AIDS Rep.* **12** (2015), 437.
- [5] C. Catalani, et al., mHealth for HIV Treatment & Prevention: A Systematic Review of the Literature, *Open AIDS J.* **7** (2013), 17–41.
- [6] M. Cabecinha, et al., Finding sexual partners online: prevalence and associations with sexual behaviour, STI diagnoses and other sexual health outcomes in the British population., *Sex. Transm. Infect.* **93** (2017).
- [7] *The European Men-Who-Have-Sex-With-Men Internet Survey*, ECDC, 2017.
- [8] P.S. Sullivan, et al., Usability and Acceptability of a Mobile Comprehensive HIV Prevention App for Men Who Have Sex With Men: A Pilot Study., *JMIR MHealth UHealth.* **5** (2017), e26.
- [9] CRH Aicken, et al., Young people's perceptions of smartphone-enabled self-testing and online care for sexually transmitted infections: qualitative interview study, *BMC Public Health.* **16** (2016)

Improving Dengue Surveillance System with Administrative Claim Data in Indonesia: Opportunities and Challenges

Atina HUSNAYAIN^{a,b}, Anis FUAD^b, Ida Safitri LAKSONO^c, Emily Chia-Yu SU^{a,d,1}

^a Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taiwan

^b Department of Biostatistics, Epidemiology, and Population Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Indonesia

^c Dr Sardjito General Hospital, Indonesia

^d Clinical Big Data Research Center, Taipei Medical University Hospital, Taiwan

Abstract. Administrative claim data is believed as one of the promising data set to augment the mandatory surveillance system which suffered from under-reporting and delay in reporting. Therefore, this study aims to examine whether the Indonesian National Health Insurance (INHI) sample data could complement dengue case-based surveillance system in a more practical way. Afterwards, this analysis also identified several future opportunities and challenges in improving the dengue surveillance system. We utilized the referral care table linked with capitation and non-capitation-based primary care service table from 2015-2016. Data cleaning, query and visualization were performed using Tableau Public and Microsoft Power BI. Result shows that dengue referral pattern is indicating the opportunity to detect dengue cases in an earlier stage and high utilization of referral care disclose the patient behaviour. Therefore, anonymous INHI sample data set potentially to complement dengue traditional surveillance system. A huge number of health facilities as data providers, bridging and interoperability chance and opportunity of early detection are identified as future opportunities. However, we also determine challenges involving how to provide the mechanism for the quick and interoperable reporting system, how to construct supportive regulation and anticipatory approach regarding the change in dengue diagnosis criteria as the implementation of ICD 11 code. Thus, practical approaches should be prepared to support the utilization of INHI sample data.

Keywords. Dengue, public health surveillance, claim data, Indonesia

1. Introduction

Dengue was identified as the most rapidly spreading vector-borne disease including in Indonesia with national incidence rate count for 24.73 per 100,000 population [1, 2]. And even has been circulated over 50 years, dengue viruses remain to persist in 440 districts or 85.60% of total area and resulting in an immense burden [2]. Therefore, an adequate dengue surveillance system might be a key to tackle the dengue occurrence.

¹ Corresponding Author, Graduate Institute of Biomedical Informatics, Taipei Medical University 15F., No. 172-1, Sec. 2, Keelung Rd., Daan Dist., Taipei, Taiwan, Email: emilysu@tmu.edu.tw.

But unfortunately, most contributor countries for global dengue infections are suffering from poor reporting systems [1] that result in a huge under-reporting issue. Research from Bandung, Indonesia, reveals that the adequacy of case reporting was only 45.7% [3]. Another problem that encounters with dengue surveillance system in Indonesia is the delay in reporting that takes days to month even though dengue cases should be reported within 24 hours after confirmed diagnosis [3, 4]. However, adequacy and timeliness are important aspects in order to provide a near-real-time response within the dengue incubation period.

Under-reporting and delay in reporting might come because of the absence of the integrated reporting system. Dengue reports from hospitals are reported to the district health offices with diverse formats through email, online messaging, or paper-based reporting system. Then, data from districts are submitted monthly to the province and national level with various structures of the report and in aggregate forms. This condition may cause dengue control, and prevention programs act too late. Thus, strengthening the dengue surveillance system requires a massive effort, including initiatives to identify potential data sources that sought the traditional surveillance system.

Given the widespread implementation of Universal Health Coverage (UHC), many studies reveal the novel uses of health claim data to augment the mandatory reporting system and provide disease-specific information [5-9]. This approach is potential to be examined in Indonesia since UHC has been implemented in 2014, along with the implementation of online claim reporting system which covers 221.203.615 members and 27.315 health facilities including the private sectors [10, 11].

Recently, as part of the open data implementation, the Indonesian Health Insurance Agency (IHIA) launched the sample data that record transactional data either in primary or secondary health care [12]. These data sets derived from 1.7 million members that explained in four relational tables, including membership, capitation, non-capitation-based primary health care and referral care table [13]. And even only derived from 1% of the total membership in 2015 using stratified random sampling [12], this data set contains a relatively short range of data records [14]; still, this data set could be the only nationally representative administration data set that is currently available and potentially updated yearly.

Accordingly, this study aims to examine whether INHI sample data could complement dengue case-based surveillance system in a more practical way. Previous studies defined the usefulness of health claim data using similarity standards with routine surveillance systems [5-7]. Nevertheless, assessing whether health insurance claim data could provide disease-specific information could be more useful for decision making. Hence, this study evaluated the utilization of INHI sample data through assessment of dengue specific information related to referral patterns. Afterwards, this analysis also identified several future opportunities and challenges in improving the dengue surveillance system using INHI sample data.

2. Method

This study uses anonymous INHI sample data (2015-2016). In order to provide dengue specific information related to referral patterns, we utilized the referral care table linked with capitation and non-capitation-based primary care service table. Dengue cases included in this analysis were in-patient cases, A91 ICD-10 code and primary visit. The patient IDs of dengue cases then were traced in capitation and non-capitation-based

primary care service table within seven days before patients were hospitalized. Severity level and health outcome then retrieved from referral care table. Data cleaning, query and visualization were performed using Tableau Public and Microsoft Power BI to illustrate dengue referral patterns. And as a future depiction, we also identified opportunities and challenges regarding the utilization of INHI sample data by comparing it with current policies on dengue surveillance system applied in Indonesia.

3. Results

Five thousand eight hundred ninety-one (5,891) dengue cases from the referral care table were obtained during 2015-2016. Within seven days, 1,131 cases (19%) were recorded in primary health care both in capitation and non-capitation-based primary care service. 13% of those cases then were registered in referral care (outpatient) along with 6% of dengue cases that treated immediately in outpatient care. But, most dengue cases (75%) were identified as having no interaction with the gatekeeper before hospitalized.

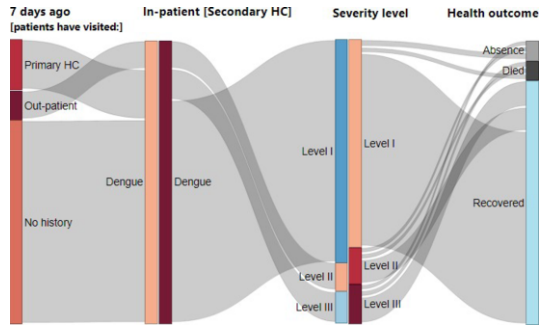


Figure 1. Referral patterns of dengue cases (2015-2016)

Furthermore, the Sankey diagram illustrated the dengue referral pattern, as shown in Figure 1. This graph showed dengue treatment stream linked with severity level and health outcome. Sankey diagram showed that most of the dengue cases were identified as having no interaction with the gatekeeper before hospitalized even 82% of cases were classified as level 1 severity.

Hereinafter, future opportunities and challenges regarding the utilization of INHI sample data are identified and listed in Table.1

Table 1. Future opportunities and challenges regarding the utilization of INHI sample data

Num	Opportunities	Challenges
1.	An increasing number of hospitals are potentially providing dengue data as their contribution to the UHC system	How to develop a mechanism for the quick and interoperable reporting system
2.	19% of dengue cases have been detected in primary health care, promising the early detection before hospitalization	How to construct regulation involving primary health care into the dengue surveillance system
3.	IHIA has providing bridging facility between the IHIA reporting system (PCare and INA CBGs) and health centre information system (Simpus) or hospital information system. Therefore, dengue cases recorded in gatekeeper and referral care can	How to anticipate change in dengue diagnosis criteria as the implementation of ICD 11 code

be easily managed cross-platform for surveillance purposes

4. Discussion

Based on the Sankey diagram provided as a result, this study successfully demonstrated that anonymous INHI sample data could provide disease-specific information related to dengue referral patterns. Thus, INHI sample data potentially complemented dengue case-based surveillance systems in a more practical way. Provided information could be used for decision making.

Observed dengue referral patterns indicate two important findings. The first one is the opportunity to detect dengue cases earlier before hospitalization as 19% of dengue cases have been detected in primary health care. This chance is supported by the bridging facility between IHIA reporting system (Primary care or PCare) and public health centre information system (Simpus). Therefore, dengue cases recorded in primary health care can easily manage cross-platform for surveillance purposes. However, the challenge is how to construct regulation that counts primary health care into the dengue reporting system.

Current regulation [15] involves gatekeeper in the system, required them to send weekly and monthly dengue cases to district health offices. But decision makings related to dengue control and prevention are taken using dengue reports provided by referral health care facilities. In fact, involving gatekeeper into the dengue surveillance system is not a big through. Since previous regulation has initiated an integrated dengue surveillance system that put gatekeeper and referral care as the same level data source [16]. Thus, strengthening policies are needed.

Another significant finding is the massive utilization of referral care discloses the patients' behaviours that tend to access the hospital. It may be caused by the rise of literacy level [17] as the knowledge of dengue symptoms, and risk perception play an important reason for visiting a doctor [18]. However, it might also an opportunity since dengue cases recorded in the IHIA reporting system (Indonesia case-based groups or INA CBGs) in hospitals also could be linked with the hospital information systems. Hence, dengue cases could be directly imported to the dengue surveillance system through data linkage. This approach would potentially shorten the dengue tiered reporting system. And as the increasing number of hospitals that contribute to the UHC system, it might be a good opportunity since UHC in Indonesia target the national coverage in 2023. Consequently, developing a mechanism for quick and interoperable reporting systems is required.

Successful implementation of a simple dengue reporting system has been recorded in Taiwan [19], which might be a good example. Minimal reporting steps lead to high adherence (99%) to report cases within 24 hours. Thus, dengue surveillance system utilizing claim data should be designed simply and facilitated with interoperability features between PCare, INA CBGs, Simpus, hospital information system, local, provincial and national health offices.

The latter action is also necessary to help anticipate change in dengue diagnosis criteria. Although ICD 11 code was released on June 2018, the INHI sample data still makes use of the former ICD 10 code. In the new ICD 11 code, dengue is divided into

four different diagnoses [20] which is completely different from the previous classification.

In conclusion, INHI sample data could provide disease-specific information related to dengue referral patterns. Therefore, this sample data set potential to complement the dengue traditional surveillance system. A huge number of health facilities as data providers, bridging and interoperability chance and opportunity of early detection are identified as future opportunities. However, we also determine several challenges involving how to provide the mechanism for the quick and interoperable reporting system, how to construct supportive regulation and anticipatory approach regarding change in dengue diagnosis criteria as the implementation of ICD 11 code. Thus, practical approaches should be prepared to support the utilization of INHI sample data.

References

- [1] A. Wilder-Smith and P. Rupali, Estimating the dengue burden in India, *The Lancet Global Health* **7**(8) (2019), e988.
- [2] Indonesian Ministry of Health, *Indonesian Health Profile [2018]*, Indonesian Ministry of Health, Jakarta, 2018.
- [3] R. Adrizain, D. Setiabudi and A. Chairulfatah, Hospital-based Surveillance: Accuracy, Adequacy, and Timeliness of Dengue Case Report in Bandung, West Java, Indonesia of 2015, *J Glob Infect Dis* **10**(4) (2018), 201-205.
- [4] *Regulation of Indonesian Ministry of Health Number 1501*, Indonesia, 2010.
- [5] J. Jung, et al., Complementing conventional infectious disease surveillance with national health insurance claims data in the Republic of Korea, *Sci Rep* **9**(1) (2019), 1-3.
- [6] E. Marder, et al., Assessment of administrative claims data for public health reporting of Salmonella in Tennessee, *J Am Med Inform Assoc* **22** (2015), e34-e37.
- [7] M.L. Lindegren, L.M.C., B. Barnes, E. Mitchel, S. Jones, W. Schaffner, Assessment of Administrative Medical Claims Data for Public Health Surveillance of Invasive Group A Streptococcal Necrotizing Fasciitis in Tennessee, *Public Health Reports* **131** (2016), 560-563.
- [8] E.L. Stockbridge, et al., Predictors of latent tuberculosis infection treatment completion in the US private sector: an analysis of administrative claims data, *BMC Public Health* **18**(1) (2018), 1-5.
- [9] S. Tanihara and S. Suzuki, Estimation of the incidence of MRSA patients: evaluation of a surveillance system using health insurance claim data, *Epidemiol Infect* **144**(11) (2016), 2260-2263.
- [10] R. Agustina, et al., Universal health coverage in Indonesia: concept, progress, and challenges, *The Lancet* **393**(10166) (2019), 75-76.
- [11] Indonesian Health Insurance Agency, *Number of UHC Members and Health Facilities*, <https://faskes.bpjs-kesehatan.go.id/aplicares/#/app/peta>, 2019.
- [12] I. Ariawan, et al., *Code Book of Indonesian Health Claim Sample Data Set [2015-2016]*, Jakarta, 2019.
- [13] A. Fuad, Indonesian Social Health Insurance Sample Data are Now Available and Accessible for Research, *Pharmaco Economics* (2019), 1.
- [14] JY. Ng, et al., National Health Insurance Databases in Indonesia, Vietnam and the Philippines, *Pharmaco Economics – Open* (2019), 1-8.
- [15] Indonesian Ministry of Health, *Dengue Control and Prevention Guideline in Indonesia*, Indonesian Ministry of Health, Jakarta, 2017.
- [16] *Regulation of Indonesian Ministry of Health Number 1479*, Indonesia, 2013.
- [17] A. Jana and R. Basu, Examining the changing health care seeking behavior in the era of health sector reforms in India: evidences from the National Sample Surveys 2004 & 2014, *Global Health Research and Policy* **2**(6) (2017), 1-4.
- [18] J. Elsinga, et al., Health Seeking Behaviour and Treatment Intentions of Dengue and Fever: A Household Survey of Children and Adults in Venezuela, *PLoS Neglected Tropical Diseases* **9**(12) (2015), 1-12.
- [19] C. McKerr, et al., Evaluation of the national Notifiable Diseases Surveillance System for dengue fever in Taiwan, 2010-2012, *PLoS Negl Trop Dis* **9**(3) (2015), 1-4.
- [20] World Health Organization, *ICD-11 for Mortality and Morbidity Statistics (ICD-11 MMS) 2018 version*, <https://icd.who.int/browse11/l-m/en>, 2019.

Intelligent Tools for Precision Public Health

Anya OKHMATOVSKAIA^{a,1} and David L. BUCKERIDGE^b
^aMcGill University

Abstract. The idea of “precision public health” (PPH) was proposed as an alternative to a one-size-fits-all approach to improving population health, which is not always effective. PPH aims to develop and apply interventions in a customized way, taking into account the detailed information about the target group. To enable the implementation of PPH in practice, we are developing an ontology-driven software platform that provides: a) access to detailed up-to-date information about population health, b) a structured machine-readable repository of evidence about public health interventions, and c) a set of intelligent tools to facilitate the assessment of evidence transferability, i.e. to determine how well certain interventions fit a given population.

Keywords. Evidence-informed decision-making, ontology

1. Introduction

Public health organizations tend to develop general policies and programs using a one-size-fits-all approach to improving the health of populations, which is not always effective in achieving the desired outcomes. Recently, the idea of “precision public health” (PPH) was proposed, where a variety of traits of subpopulations or even communities are detected and tracked to enable the development and targeted application of interventions that are best suited to that group [1]. This concept of ‘precision’ is borrowed from the more familiar context of ‘precision medicine’, which refers to targeting therapy (particularly for cancer) using detailed data about individual genomic and other traits on one hand, and available evidence on the effectiveness of treatment options in different patients under different conditions, on the other hand. At a population scale, PPH similarly aims to use the detailed information about a defined population and evidence about the effectiveness of available interventions in different populations under different conditions, to precisely match an intervention to a population for best results.

Despite the general appeal of this vision, the computational tools to support the implementation of PPH in practice are still lacking. Most PPH frameworks agree on prioritizing the development efforts in the three areas: 1) access to detailed population health information, 2) access to evidence on health interventions, and 3) tools for combining population information and evidence. Artificial intelligence methods, including semantic technologies, have the potential to address the challenges in all three priority areas of PPH. Using software ontologies to organize both population health information and intervention evidence can facilitate access to these resources and support the development of intelligent decision-support tools. In the rest of the paper, we first introduce our initial work in enhancing access to population health information through the Population Health Record (PopHR) project, we then describe the Interventions Knowledgebase, which extends PopHR to provide access to evidence about public health

¹ Anya Okhmatovskaia, Corresponding author, E-mail: anna.okhmatovskaia@mcgill.ca.

interventions, and finally, we discuss future work towards the development of intelligent tools for transferability assessment.

2. Population Health Record

Just like personalizing therapy in precision medicine, developing targeted public health interventions requires the capacity to integrate, analyze, and interpret a growing volume of heterogeneous population health data. Decision-makers must have access to timely and accurate indicators of population health, including health information about diseases and risk factors and information about available resources. Regrettably, state-of-the-art health information systems tend to suffer from certain weaknesses in their design and operation, including out-of-date static indicators, insufficient spatio-temporal resolution, and limited transparency of computational algorithms and data sources. As a result, they do not provide sufficient precision in describing populations that could be targets for interventions. The PopHR is an ontology-driven web-based software platform that was designed to address the shortcomings of the existing systems and provides access to comprehensive, timely and detailed information about population health [2].

In order to collect, align, integrate, and transform data from multiple sources, the PopHR relies on a suite of ontologies that encode an explicit semantic framework for population health. This framework includes taxonomies of population characteristics, diseases, determinants of health, a multi-axial classification of health indicators, spatio-temporal relations, etc. In addition, it incorporates epidemiological causal knowledge, identifying upstream and downstream relationships for different health conditions. The PopHR uses this framework to organize and present health indicators in a meaningful way. The current version of the PopHR focuses on chronic diseases and their risk factors. With partners from the Institut National de Santé Publique du Québec (INSPQ) and funding from the Public Health Agency of Canada (PHAC), we are in the process of deploying the PopHR in the province of Quebec for routine use by public health professionals.

By providing access to the detailed population health information, PopHR supports one of the PPH priority areas, and serves as a foundation for the development of intelligent tools in the other two areas. Extending the PopHR to support integrated access to evidence about interventions and decision-support tools that would suggest to the user the interventions that are likely to be the most effective in a given population (something that requires juxtaposition of evidence about an intervention with information about a population in a target region) is the focus of our current research. As far as we know, no existing system currently supports such functionality.

3. Interventions knowledgebase

Biomedical literature has accumulated an enormous amount of evidence documenting the effectiveness of health interventions in many populations and settings. This evidence, however, is neither readily accessible, nor encoded consistently, and therefore not easy to use for the purpose of PPH. The field of knowledge translation (KT) has made considerable progress to address this problem [3], including the development of searchable online repositories of evidence [4]. To date, however, these repositories rely heavily on manual curation, with the associated high costs of building and maintaining them. Keeping

up with the exponentially growing body of biomedical evidence comes at a cost of reduced precision of these repositories: most of them limit their focus to systematic reviews and exclude the original studies, and annotate the evidence coarsely with only a small set of intervention and study design characteristics, insufficient for precise matching of the evidence to target populations.

To enable the encoding of evidence with a precision sufficient for assessing its transferability to a defined population in the context of PPH, one needs a sufficiently expressive vocabulary. While all existing web repositories for health evidence use some sort of an underlying conceptual model to describe interventions and their effects, the model is often shallow and informal, which limits precision and presents a barrier for the use of automated tools to process such evidence. In contrast, a formal intervention evidence framework in the form of an ontology will provide the necessary level of expressivity, enable logical reasoning over existing evidence, and provide a foundation for automated retrieval and annotation of new evidence.

The most popular standard framework for describing health evidence, PICO, has been formalized as an ontology by the Cochrane group [5] and it identifies the essential components of evidence (population, intervention, comparison, outcome) and the relationships among them. The high-level PICO framework can be supplemented by more detailed vocabularies to further classify instances within each component. However, while extensive terminologies exist to systematize clinical interventions, and behavior change techniques, little work has been done to develop a standard vocabulary for describing public health interventions. To address this gap, we have developed an Ontology of Public Health Interventions (OPHI), which can be used to formally encode evidence about intervention effectiveness in a standard, machine-readable format.

3.1. Ontology of Public Health Interventions (OPHI)

To ensure that our ontology reflects a consensus among public health professionals regarding intervention classes and properties, we conducted a scoping review of literature to identify existing conceptual frameworks and classifications of public health activities. The search extended over articles in public databases (PubMed, Scopus, Google Scholar and Cochrane Library) as well as the grey literature: documents from PHAC, INSPQ, Public Health Europe, WHO, and the National Public Health Partnership. Two independent researchers examined the titles and abstracts of nearly 1000 articles for relevance, leaving around 200 articles for full-text review, yielding a total of fourteen sources to be synthesized. These sources included several formal ontologies and classification systems, as well as informally described conceptual frameworks and diagrams.

We extracted concepts from each of the 14 proposed frameworks along with their definitions. Some of the sources did not formally define any concepts that were used. As a result, many of the concept definitions had to be inferred based on the provided examples of usage and lists of child concepts or property values. We also identified a number of concepts that were used synonymously across different frameworks, as well as ambiguous terms that had different meanings in different frameworks. After the required disambiguation and merging of synonymous terms, we developed a synthetic conceptual model and formalized it using Web Ontology Language (OWL).

The resulting Ontology of Public Health Interventions (OPHI) builds on the basic structure introduced by the PICO ontology to capture health evidence, augments it with

the concepts and relationships from the synthesized frameworks, and provides a multiaxial classification of interventions based on settings, recipients, purpose and methods. As part of the PopHR ontology suite, OPHI is directly linked to other ontologies we developed previously for this platform, notably, the ontology of public health (PHOnt), which encodes a broader range of concepts relevant to public health practice (e.g. public health resources), and includes causal knowledge about health conditions and health determinants. This linkage is essential when it comes to assessing the applicability and feasibility of implementing interventions in new contexts and predicting their effects in target populations. Although OPHI specifically focuses on public health interventions, the underlying knowledge model is compatible with and suitable for encoding clinical interventions as well. This is important since evidence from clinical trials is more abundant compared to studies evaluating the effects of public health programs.

3.2. Sample knowledgebase of influenza interventions

To demonstrate how OPHI can be used within the PopHR infrastructure to support PPH practice, and to verify its suitability for annotating studies of interventions, we have created a small sample knowledgebase of interventions, focusing on influenza control². As much of the evidence related to influenza control focuses on vaccination, we have extended OPHI with a well-established Vaccine Ontology [6] to increase the annotation precision in this area. We selected four systematic reviews that focused on diverse aspects of influenza control measures (vaccine safety, effectiveness, vaccine uptake, non-vaccination prophylactic measures) and retrieved fifty-six associated studies. We manually extracted from this corpus the essential findings about the effectiveness of interventions and annotated all relevant terms and features of the interventions and study populations. We were able to align the relevant properties of study design, categories of interventions, characteristics of study population and types of outcome with the existing OPHI terms. Figure 1 presents a fragment of the encoding.

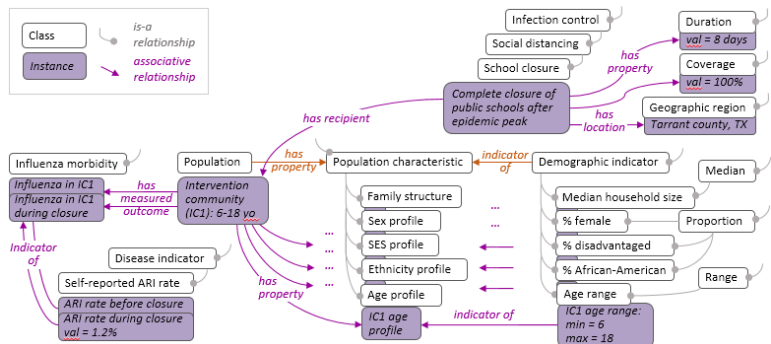


Figure 1. Sample encoding of influenza intervention evidence using OPHI (fragment)

In our pilot work building a sample knowledgebase of influenza control interventions, we demonstrated that OPHI provides a suitable semantic framework for describing and classifying existing evidence about public health interventions. While we believe that our ontology can be applied successfully to other subject areas, our experience encoding

² Subject area has been chosen because of the high burden associated with influenza, and because it has been extensively studied within clinical trials, and public health programs, and large amounts of evidence is available.

influenza vaccination studies reveal the potential need to incorporate additional vocabularies to achieve the necessary level of precision. Combined with the causal knowledge already captured in PopHR ontologies, the OPHI-based interventions knowledgebase would support searches for interventions with specific properties and with demonstrated effectiveness in specific populations and settings and will provide a foundation for the development of the intelligent tools for evidence transferability assessment. Another direction of our future work is exploring the potential of employing statistical natural language processing (NLP) techniques to automatically populate the knowledgebase with instances of interventions and evaluation findings from published literature and databases. Such automation has been previously demonstrated as feasible in the context of clinical trials [7]. The influenza interventions knowledgebase will provide a set of annotations that can be used for training the machine learning algorithms.

4. Future work

The third PPH priority area has seen the least amount of progress. In the context of PPH, to make an evidence-informed decision, one must have not only detailed information about target population and evidence of interventions' effectiveness but also the cognitive capacity to assess the 'fit' of the evidence to the target population. Systematic reviews, which constitute the standard of evidence for public health decision makers, estimate an intervention's impact over a range of settings and provide limited detail on the populations and conditions of included studies. This means that extensive additional information gathering from individual studies is usually needed to assess the potential effects in a local population setting. Given the complexity of this task, decision-support tools have the potential to facilitate this process significantly. To date, however, there has been little research on the development of tools and software to help practitioners effectively combine evidence about health interventions with population information.

The OPHI-based interventions knowledgebase will make characteristics of the study population used to generate evidence readily available for access by automated tools, while corresponding characteristics of the target population can be retrieved from PopHR data. With a set of conversion rules that account for various measures of the same characteristic, it will be possible to compare the two populations using measures of statistical and semantic similarity [8], and assess how well an intervention "fits" a given population and possibly estimate the expected effect of that intervention.

References

- [1] S. Desmond-Hellmann, Progress lies in precision, *Science* **353** (2016), 731–1.
- [2] A. Shaban-Nejad, M. Lavigne, A. Okhmatovskaia, D.L. Buckeridge, PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data, *ANYAS*, (2016), 44–53.
- [3] S.E. Straus, J. Tetroe, I. Graham, Defining knowledge translation. *CMAJ*, **181**(3–4) (2009), 165–8.
- [4] M. Dobbins, K. DeCorby, P. Robeson, H. Husson, D. Tirilis, L. Greco, A knowledge management tool for public health: health-evidence.ca, *BMC Public Health*, **10**(1) (2010), 496.
- [5] Cochrane Collaboration, PICO Ontology (2016) <https://linkeddata.cochrane.org/pico-ontology>
- [6] Y. He, L. Cowell, A. D. Diehl et al., VO: vaccine ontology, in *Proceedings of the 1st International Conference on Biomedical Ontology (ICBO '09)*, Buffalo, NY, USA, (2009).
- [7] B.C. Wallace, J. Kuiper, A. Sharma, M.B. Zhu, I.J. Marshall, Extracting PICO Sentences from Clinical Trial Reports using Supervised Distant Supervision, *Journal of Machine Learning Research*, **17** (2016).

- [8] B. Gallego, S.R. Walter, R.O. Day, A.G. Dunn, V. Sivaraman, N. Shah, et al. Bringing cohort studies to the bedside: framework for a “green button” to support clinical decision-making. *J Comp Eff Res*. Future Medicine Ltd London, UK, **4(3)** (2015), 1–7.

Machine Learning-Based Identification of Obesity from Positive and Unlabelled Electronic Health Records

Vicent BLANES-SELVA ^{a,1}, Salvador TORTAJADA ^b Ruth VILAR ^c

Bernardo VALDIVIESO ^c Juan M. GARCÍA-GÓMEZ ^a

^a *Instituto Universitario de Tecnologías de la Información y Comunicaciones. Universitat Politècnica de València. Camino de Vera s/n. 46022 Valencia, España*

^b *Instituto de Física Corpuscular (IFIC), Universitat de València, Consejo Superior de Investigaciones Científicas, 46980 Paterna, Spain*

^c *Unidad Mixta de Tic aplicadas a la reingeniería de procesos socio-sanitarios ERPSS, Instituto de Investigación Sanitaria La Fe, Avenida Fernando Abril Martorell, 46026 Valencia, Spain*

Abstract. Introduction: Prevalence of overweight and obesity are increasing in the last decades, and with them, diseases and health conditions such as diabetes, hypertension or cardiovascular diseases. However, hospital databases usually do not record such conditions in adults, neither anthropomorphic measures that facilitate their identification.

Methods: We implemented a machine learning method based on PU (Positive and Unlabelled) Learning to identify obese patients without a diagnose code of obesity in the health records.

Results: The algorithm presented a high sensitivity (98%) and predicted that around 18% of the patients without a diagnosis were obese. This result is consistent with the report of the WHO.

Keywords. Obesity, Overweight, Identification, Machine Learning, PU Learning

1. Introduction

Overweight and obesity epidemic prevalence has increased since the 1970s in most of the countries, affecting children and adults alike [1,2]. According to the World Health Organization (WHO) in Europe, more than 50% of the population is overweight, and 20% is obese (WHO Data and statistics, 2019). In Spain, obesity already affects 17.4% of the adult population, whereas considering obesity and overweight together, more than half (54.5%) of adults are overweight.

The most challenging feature of this situation is the increasing trend of people with obesity and overweight in our society. This trend is especially worrying due

¹Corresponding author. Email: viblasel@upv.es

to conditions and diseases that are associated with an excessive weight such as diabetes, hypertension or cardiovascular disease [3].

However, this increase in obesity and overweight among the population has not been reflected in hospital practice, which usually attends the problems derived from obesity but not the underlying condition [4]. For example, at the Hospital Universitario La Fe, in Valencia (Spain), the percentage of patients in the electronic health records with a diagnosis of obesity (278.00 and 278.01 ICD9 codes) is 2.4%. Also, the lack of anthropometric measures in the databases (only around 3%) makes it difficult to extract useful indicators for identification, such as the Body Mass Index (BMI).

Public policies and hospital interventions based on that data may lack correctness, producing biases in the conclusions extracted from it. Under this situation, other methods to know the real condition of patients are required. Among others to cross-reference with primary care data, active questionnaires to patients or data imputation are actions to mitigate this limitation.

To solve this identification issue, several restrictions should be considered 1) No specific information is compiled about obesity and overweight in the adult population; 2) Majority of samples are not labelled as obesity, overweight or no obesity, neither BMI is collected 3) real classes are unbalanced given the statistics about WHO.

With this scenario, we propose to develop a machine-learning model based on positive-unlabelled (PU) learning for obesity identification trained by positive and unlabelled electronic health records as a public health tool to estimate hidden positive patients with obesity before studying comorbidities at patient aggregation level. The authors developed this work in the context of the H2020 European project CrowdHealth (<https://crowdhealth.eu/>) to provide helpful tools for public policies' developers.

2. Materials and Methods

In this study, we analysed data from a total of 49805 anonymised patients represented by their basic demographics and information coming from the Electronic Health Records of HULAFE such as external specialist consultations, emergency room registries, hospital admissions, laboratory results and home hospitalization (UHD) registries. Ethical approval was obtained for this study from the Ethical Committee of Hospital Universitario La Fe (Ethical Code: 2016/0350).

We searched for conditions associated with obesity and overweight in the scientific literature [2,3,5] and created a list of concepts to query against the data, extracting a final set of predictive variables. These variables were mostly related to the frequency of diagnoses, frequency of visits to specialist and laboratory results about cholesterol. A total of 32 variables (see Table 1) composed the final version of the dataset. From the 49805 patients that formed our sample, 7033 (~ 14.12%) were labelled as positive for obesity, having an ICD9 code for obesity present, and 42772 (~ 85.88%) were unlabelled.

Table 1. Variables obtained from the dataset

Variables		
Stroke	Angiologists consults	Circulatory diags
Hypertension diags	LDL analyses	HDL analyses
Triglycerides analyses	Endocrine diags	Insuline Tests
Insuline-related emergencies	Cardiologist consults	Infarction diags
Cardiology diags	Coronary-related admission	Gastric issues
Gallbladder-related emergencies	Gallbladder-related admissions	Rheumatology consults
Musculoskeletal diags	Sleep issues	Sleep breathing problems
UHD onco	Radio or Quimio events	Oncologist visits
Protein analysis	Psiquiatric diags	Pain diags
LDL mean	HDL mean	Triglycerides mean
Diabetes diags	Age	

As mentioned before, the number of identified obesity patients in the Electronic Health Records (EHRs) are meagre. The authors hypothesise that among unlabelled patients there is a significant amount of obese patients ($\sim 17\%$ of the population) according to the statistics by the WHO. This problem belongs to the class imbalance type, which is characterised for having a majoritarian class where most of the samples belong [6].

Therefore a classic machine learning classification algorithm is not appropriate for this task. As a consequence, a correct approach under machine-learning discipline for dealing with our problem is PU Learning. Liu et al. in [7] define PU Learning as a semi-supervised technique for building a binary classifier based on positive and unlabelled examples. In PU Learning, two sets of samples are available for training: the P set of positive instances, and a U set, assumed to contain a mixture of both positive and negative examples but unlabeled at training time. This technique is particularly useful when the level of contamination, positive samples among negatives, in U is low (around 15%).

In concrete, we have used an implementation of the algorithm "Bagging Inductive PU Learning" (see algorithm 1) proposed by Mordelet et al. [8]. In this version, an ensemble of T classifiers is trained using the P set and a randoms sub-sampling of U set with size $|P|$ for each classifier. We implemented Mordelet's algorithm in Python 3 [9], using instances of the Random Forest (RF) [10] from the *scikit-learn* library [11] as classifiers in the ensemble.

To demonstrate the validity of PU Learning in the obesity identification problem we developed an experiment where we split the data randomly, but keeping the class proportion, in two sets: train (80%) and test (20%). After that we trained an ensemble of 11 models using Mordelet's algorithm and evaluated the sensitivity and the amount of negatives in U . This last metric is calculated as the specificity but taking into consideration that does not exist a real ground truth for unlabelled/negatives. We repeated 100 times the same random experiment to get different partitions of train and test and reported the mean and 95% confidence interval of the metrics.

Algorithm 1 Mordelet's Inductive bagging PU learning**INPUT:** P ; U ; K = size of bootstrap samples, T = number of bootstraps**OUTPUT** a function $f : \chi \rightarrow \mathbb{R}$

```

1: for  $t = 1$  to  $T$  do
2:   Draw a subsample  $U_t$  of size  $K$  from  $U$ .
3:   Train a classifier  $f_t$  to discriminate  $P$  against  $U_t$ .
4: end for
5: return  $\frac{1}{T} \sum_{t=1}^T f_t$ 

```

3. Results

The results of the experiment are summarised in Table 2. The algorithm can identify almost every patient with obesity, obtaining a sensitivity of 98%. Also, it estimates that the U contains around 18% of persons with an obesity condition, which is similar to the estimations reported by the WHO.

Table 2. Experimental results over 100 iterations using PU Learning

Metric	Mean	Confidence Interval 95%
Sensitivity	0.981	0.981, 0.982
Estimated negatives in U	0.822	0.820, 0.823
Estimated positives in U	0.178	0.177, 0.180

4. Discussion

As seen in the previous section, almost every obesity case presented to the algorithm is recognised correctly, presenting a 98.1% of sensitivity. The amount of negatives in U identified by the algorithm is around 82%, which means that the 18% of the patients in U are obese. These numbers are coherent with the statistics reported by the WHO. The results are relevant because, aside from being promising, no negative samples are required to train the model and the result was obtained with a relatively large N (49805).

The main contribution of this work consists in the application of one known machine learning algorithm to the identification of patients with obesity using a small fraction of labelled data. This solution could be useful to identify obesity at the patient level and mitigate the under-representation of this condition among the EHR [4]. This algorithm could be used to implement an automatic tool used in workflow mining, comorbidities statistics and screening, providing hospitals with a rapid method for detecting patients target for a program or intervention related to obesity.

The main weakness of this method is its bias to ignore obese patients that have made little use of the hospital resources. This effect is due to the variables

used to make the models, which are based mostly on the frequency of visits to specialist and diagnosis.

As future work, the most straightforward way of profiting this model is its implementation as part of a web-based platform, allowing users to send their data to get the predictions from the model and some statistics at the end of the interaction.

5. Conclusion

The implementation of the PU Learning methodology in the obesity identification produced satisfactory results, providing high sensitivity and identifying a similar number of patients with obesity to the WHO data. This model can be used to identify patients with obesity among the hospital's population to prepare interventions or public policies.

References

- [1] Robert J Kuczmarski, Katherine M Flegal, Stephen M Campbell, and Clifford L Johnson. Increasing prevalence of overweight among us adults: the national health and nutrition examination surveys, 1960 to 1991. *Jama*, 272(3):205–211, 1994.
- [2] Boyd A Swinburn, Gary Sacks, Kevin D Hall, Klim McPherson, Diane T Finegood, Marjory L Moodie, and Steven L Gortmaker. The global obesity pandemic: shaped by global drivers and local environments. *The Lancet*, 378(9793):804–814, 2011.
- [3] George A Bray. Medical consequences of obesity. *The Journal of Clinical Endocrinology & Metabolism*, 89(6):2583–2589, 2004.
- [4] J Cleator, E Richman, KS Leong, L Mawdsley, S White, and J Wilding. Obesity: under-diagnosed and under-treated in hospital outpatient departments. *International journal of obesity*, 26(4):581, 2002.
- [5] MF Rolland-Cachera, M Deheeger, M Maillot, and F Bellisle. Early adiposity rebound: causes and consequences for obesity in children and adults. *International journal of obesity*, 30(S4):S11, 2006.
- [6] Nathalie Japkowicz. The class imbalance problem: Significance and strategies. In *Proc. of the Int'l Conf. on Artificial Intelligence*, 2000.
- [7] Bing Liu, Wee Sun Lee, Philip S Yu, and Xiaoli Li. Partially supervised classification of text documents. In *ICML*, volume 2, pages 387–394. Citeseer, 2002.
- [8] Fantine Mordelet and J-P Vert. A bagging svm to learn from positive and unlabeled examples. *Pattern Recognition Letters*, 37:201–209, 2014.
- [9] G van Rossum. Python tutorial, technical report cs-r9526. *Centrum voor Wiskunde en Informatica (CWI)*, Amsterdam, 1995.
- [10] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

Medical Informatics in the Digital Personalized Health and Medicine Era: A SWOT Analysis and Actionable Strategies

Rada HUSSEIN^{a,1}

^a *Ludwig Boltzmann Institute for Digital Health and Prevention, Austria*

Abstract. With the rapid proliferation of digital health technologies, the role of medical informatics in advancing personalized healthcare and wellbeing is emphasized. To examine the readiness of medical informatics for fostering digital health moving forward, a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis was performed. As a result, actionable strategies have been identified to maximize the benefits of medical informatics and overcoming the faced challenges in the digital health era.

Keywords. Digital health, medical informatics, strategic planning, SWOT

1. Introduction

The movement from "eHealth" to "Digital Health" emphasizes on providing people-centred and integrated health services utilizing the technological advances, such as mobile health (mHealth), Internet of Things (IoTs), Artificial Intelligence (AI), big data analytics, cloud computing, blockchain, etc. [1].

Medical Informatics or Biomedical and Health Informatics (BMHI) has become an established scientific-multidisciplinary discipline involving computer, cognitive and social sciences [2]. BMHI utilizes health information technologies to provide the optimal use of information for improving individual health, healthcare, public health and biomedical research.

The goal of this paper is to analyze the strengths, weaknesses, opportunities and threats (SWOT) of the current status of the BMHI field. Consequently, actionable strategies are identified to leverage the anticipated potential of BMHI for the Digital Health (DH) era.

2. SWOT analysis and actionable strategies

SWOT analysis is one of the world's most widely used methods for strategic planning [3]. This paper uses this instrument to match the strengths and weaknesses of the BMHI field with opportunities and threats of DH. The SWOT analysis has been performed

¹ Corresponding Author, Rada Hussein, E-mail: rada.hussein@dhp.lbg.ac.at.

based on the review of literature, industry reports and initiatives of the global health agencies.

To identify the potential tactical strategies for maximizing the benefits of BMHI for the DH era, an extended "variant" version of the SWOT-named the Threats, Opportunities, Weaknesses and Strengths (TOWS) analysis was conducted [3]. Figure 1 depicts how to develop actionable strategies through the TOWS analysis.



Figure 1 TOWS analysis and actionable strategies

3. Medical informatics strengths and limitations

The BMHI domain has a broad scope with the full range of knowledge (research) and practice (applications) [4] covering several sub-disciplines, mainly: bioinformatics, imaging informatics, clinical informatics and public health informatics. Additionally, there are two rapidly growing domains within the BMHI field, namely: 1) Consumer Health Informatics (CHI) that addresses consumers' articulated needs with managing their health concerns [4,5], and 2) Global Health Informatics (GHI) that focuses on empowering people to use appropriate technology, aligning with a global perspective of healthcare provision to all [6].

The main limitations of scaling up the eHealth initiatives and reaching the desired impact are [6]: 1) limited interoperability with legacy systems at national scale, 2) lack of evidence regarding the eHealth impact on health system performance and individual health, 3) lack of regulatory framework concerning data privacy and security, and 4) the insufficient use of large sets of data, i.e., converting data into health intelligence. On the other hand, the main barrier that hinders CHI [4,5] is the misalignment between the used technology and the users' socio-technical environment. This can significantly impact the technology utilization, endorsement by healthcare providers and users' empowerment.

4. Digital personalized health and medicine opportunities and challenges

The DH domain is considered as a key enabler of the personalized, preventive, predictive, participative, and precision (5P) medicine through leveraging the vision of providing equitable and person-centred care [5]. Consequently, DH has been attracting great interest of all stakeholders, including academic researchers, consumers, global health agencies, healthcare industry, policymakers and others. Recent studies covered different aspects of the DH domain, for example, industry and academic reviews of the DH technical trends [1], the World Health Organization (WHO) 's practical guide for monitoring and evaluating digital health interventions as well as the draft version of the global strategy on digital health 2020-2024 [7], Stanford medicine's

report on the democratization of healthcare [8], Johns Hopkins' roadmap for digital health validation [9], and a framework for assessing the impact of connected health services [10]. At the regulatory level, the Integrating Healthcare Enterprise (IHE)-Europe published its perspective in understanding and adopting the General Data Protection Regulations (GDPR) [11]. Additionally, a vision of how the GDPR supports healthcare transformation is described in [12].

Nevertheless, the main challenge facing the DH market [5,13] is that it is like "wild west" because most of the applications are neither based on the evidence of effectiveness nor adopting existing technical standards. Furthermore, most of the DH applications are mainly technology-driven as "one-size fits all" that may lead to harmful use. At the integration level [5,13], there is a lack of interoperability between user-generated data and health records in addition to an inadequate or fragmented legal frameworks, especially in the area of privacy and data security. Moreover, there is a mandate to address the DH from the socio-technical and organizational perspectives [1] in the context of change management, user engagement, information governance, workforce capacity, citizen empowerment, regulations, business models, etc.

5. Medical Informatics and digital health TOWS analysis

Table 1 summarizes the results of the SWOT analysis and their translation into TOWS analysis and actionable strategies.

6. Medical Informatics and digital health actionable strategies

As listed in Table 1, the actionable strategies can be categorized as follows:

The SO actionable strategies aim to *shape the BMHI research and education* for the DH era utilizing the advances in science and technology. This also matches with the proposed research agenda described in [2].

The WO actionable strategies aim to *build interdisciplinary partnerships and collaborations* to tackle the pressing challenges in aligning the BMHI field with future trends in medicine and DH technologies.

The ST actionable strategies aim to *leverage the role of the BMHI associations in building the ecosystem for DH connected health* with all involved stakeholders, including researchers, consumers, global health agencies, DH industry, standard development organizations, policymakers and others.

The WT actionable strategies aim to *roadmap an integrated future for BMHI and DH* at technical, environmental and organizational levels. This requires interdisciplinary teamwork of healthcare providers, informaticians, data scientists, systems developers, consumers, and experts in behavioural science, law, human-computer interaction, business modelling, and change management.

Table 1. BMHI and DH TOWS analysis and corresponding actionable strategies

	BMHI Strengths (S) S1 Breadth and diversity of BMHI subcategories S2 Wide scale of BMHI methods, tools and applications S3 Consumer Health Informatics articulates the user-centric approach S4 Global Health Informatics focuses on improving healthcare systems and outcomes	BMHI Weaknesses (W) W1 Too many overlapping standards hinder interoperability W2 misalignment between the used technology and the "real world" socio-technical environment W3 Limited large-scale evidence W4 Limited convergence of data into health intelligence at national and global scales
DH Opportunities (O) O1 Wide scale of technological innovation in big data analytics, cloud computing, AI, wearable devices, IoT, blockchain, etc. O2 Future trends in medicine: personalized, pervasive, predictive, preventive, participatory, precise, preemptive, and patient-centralized (8P medicine) O3 The WHO global health initiatives for digital health, universal health coverage, democratization of data, evidence for all, sustainable development goals, etc. O4 Great interest of the BMHI community (academia, industry and associations) worldwide in digital healthcare transformation	SO actionable strategies: SO1 Strengthen the BMHI discipline with new trends in medicine and technology (S1, O1, O2) SO2 Establish strong partnerships for international collaborative research programs and initiatives (S1-S4, O1-O4) SO3 Align the BMHI education and research programs with the WHO global health initiatives (S3, S4, O3) SO4 Leverage the strengths of the BMHI national and international associations in developing and updating the BMHI competencies frameworks (in regular basis)- that are intended for curriculum development, accreditation and workforce certification (S1, O4)	WO actionable strategies: WO1 Take the initiative (by the international standards organizations) to coordinate their efforts in improving the existing BMHI standards and resolving the interoperability gaps (W1, O1, O2, O4) WO2 Adopt socio-technical systems theory to support a better understanding of end-user performance (W2, O1, O2, O4) WO3 Adopt the WHO guidelines on DH interventions and corresponding frameworks for DH monitoring and evaluation (W3, O1-O4) WO4 Establish international scale research programs (that adopt new methodological or technical approaches) to advance interdisciplinary understanding (W4, O1-O4)
DH Threats (T): T1 Rapid pace of technology-driven applications (as one size fits all) T2 Lack of international regulatory frameworks for DH testing, validation, safety, and data privacy and security T3 Lack of interoperability standards for user-generated data T4 Low acceptance and adoption of DH	ST actionable strategies: <i>Leverage the role of the BMHI associations to:</i> ST1 Engage all stakeholders in designing national DH connected health services and the corresponding consumers' empowerment programs (S1, S2, T1, T4) ST2 Establish multidisciplinary taskforce to validate the current initiatives of DH regulatory frameworks in addition to technical working-groups to resolve the interoperability gaps with standards development organizations (S1-S4, T2, T3) ST3 Develop a specialized competencies framework for implementing and monitoring the DH regulatory framework (S1,T2) ST4 Develop a collaborative online platform for sharing best practices and lessons learned in overcoming the DH challenges across the associations' members (S1-S4, T2, T3, T4)	WT actionable strategies: <i>Overcome weaknesses by making them strengths (move toward SO strategy)</i> <i>Leverage the role of the BMHI associations to:</i> WT1 Develop a specialized competencies framework for DH open data and science (W1, W3, W4, T1, T3) WT2 Build a multidimensional collaborative team to develop a roadmap for interdisciplinary understanding of the DH socio-technical and organizational environment (W2, T2, T4) WT3 Develop a collaborative online platform for open innovation to tackle the DH challenges (W3, W4, T1, T4) WT4 Utilize change management concepts to 1) tackle the unpredictable and complex interactions between the people and DH technology, 2) strengthen user engagement (W2, T2, T4)

7. Conclusion

Medical informatics is expected to play a pivotal role in the transformation of digital "personalized" healthcare and medicine. To achieve this goal, the described actionable strategies support paving the way of the BMHI field towards the DH era, through 1) enriching the BMHI research and education agenda with emerging DH technologies, WHO initiatives and medicine trends, 2) strengthening the role of the BMHI associations in building the ecosystem of digitally connected health, and 3) road-mapping an integrated interdisciplinary future of BMHI and DH at socio-technical and organizational levels. Thus, realizing the BMHI aim in advancing science and improving healthcare outcomes.

References

- [1] L. Rooney, S. Rimpiläinen, C. Morrison, and S.L. Nielsen, *Review of Emerging Trends in Digital Health and Care: A report by the Digital Health and Care Institute*, Digital Health & Care Institute, Glasgow, 2019. <https://doi.org/10.17868/67860>.
- [2] R. Haux, C. Kulikowski, S. Bakken, et al., Research Strategies for Biomedical and Health Informatics: Some Thought-provoking and Critical Proposals to Encourage Scientific Debate on the Nature of Good Research in Medical Informatics, *Methods of Information in Medicine*. **56** (2017), e1–e10. doi:10.3414/ME16-01-0125.
- [3] T. Sammut-Bonnici, and D. Galea, SWOT Analysis, in: C.L. Cooper (Ed.), *Wiley Encyclopedia of Management*, John Wiley & Sons, Ltd, Chichester, UK, 2015. doi:10.1002/9781118785317.weom120103.
- [4] A. Faiola, and R.J. Holden, Consumer Health Informatics: Empowering Healthy-Living-Seekers Through mHealth, *Progress in Cardiovascular Diseases*. **59** (2017), 479–486. doi:10.1016/j.pcad.2016.12.006.
- [5] M. Edmunds, C. Hass, and E. Holve, eds., *Consumer Informatics and Digital Health: Solutions for Health and Health Care*, Springer International Publishing, 2019.
- [6] L.A.G. Celi, H.S.F. Fraser, V. Nikore, J.S. Osorio, and K. Paik, eds., *Global health informatics: principles of ehealth and mhealth to improve quality of care*, The MIT Press, Cambridge, Massachusetts, 2017.
- [7] WHO Digital health, (n.d.), <https://www.who.int/westernpacific/health-topics/digital-health> (accessed October 11, 2019).
- [8] A.B. Bach, *Stanford Medicine Health Trends Report examines opportunity to democratize health care, Scope*, <https://scopeblog.stanford.edu/2018/12/13/stanford-medicine-health-trends-report-examines-opportunity-to-democratize-health-care/>, 2018.
- [9] S.C. Mathews, M.J. McShea, C.L. Hanley, A. Ravitz, A.B. Labrique, and A.B. Cohen, Digital health: a path to validation, *Npj Digital Medicine*. **2** (2019), 38. doi:10.1038/s41746-019-0111-3.
- [10] I. Chouvarda, C. Maramis, K. Livitkaia, V. Trajkovik, S. Burmaoglu, et al., Connected Health Services: Framework for an Impact Assessment, *J Med Internet Res*. **21** (2019). doi:10.2196/14005.
- [11] *IHE Perspective on the European Union GDPR | IHE Europe*, (n.d.), <https://ihe-europe.net/node/186> (accessed October 11, 2019).
- [12] B. Bernd and R. Pekka, How Does GDPR Support Healthcare Transformation to 5P Medicine?, *Studies in Health Technology and Informatics* (2019), 1135–1139. doi:10.3233/SHTI190403.
- [13] K. Huckvale, C.J. Wang, A. Majeed, and J. Car, Digital health at fifteen: more human (more needed), *BMC Med*. **17** (2019), 62. doi:10.1186/s12916-019-1302-0.

Mining Potential Effects of HUMIRA in Twitter Posts Through Relational Similarity

Keyuan JIANG ^{a,1}, Shichao FENG ^b, Liyuan HUANG ^a, Tingyu CHEN ^a and Gordon R. BERNARD ^c

^a *Purdue University Northwest, Hammond, Indiana, U.S.A.*

^b *University of North Texas, Denton, TX 76203, U.S.A.*

^c *Vanderbilt University, Nashville, Tennessee, U.S.A.*

Abstract. HUMIRA, a biologic therapy, has been approved to treat autoimmune diseases and been marketed in many countries worldwide. Much like other medications, it demonstrates many effects on the human body. It is important to understand its effects from the information generated by its users, and social media is one of the venues its users share their experience with the medication. To understand what HUMIRA effects were reported on Twitter, we utilized a relational similarity-based approach to infer HUMIRA effects based upon known medication-effect relations of other medications. With a corpus of 3.6 million preprocessed, “clean” tweets, a total of 55 effects were identified, and among them, 46 were previously observed, and nine were potentially unreported after verification with six reliable sources. The results not only indicate that many HUMIRA effects shared by the Twitter users are consistent with those previously reported, but also demonstrate the power and utility of our method, making it applicable to studying effects of other medications shared by Twitter users.

Keywords. HUMIRA, medication effects, mining social media, relational similarity, relation extraction

1. Introduction

As a blockbuster medication of AbbVie Inc., HUMIRA (adalimumab), a biologic therapy administered via subcutaneous injection, has been approved to treat autoimmune diseases in North America and the European Union as well as being approved in Japan for the treatment of intestinal Behcet’s disease. It has been sold in many countries worldwide, including Japan, China, Brazil and Australia [1]. Like other prescription medications, HUMIRA demonstrates various effects to those who took it, and its adverse effects such as serious infections and a rare type of cancer have been observed and reported [2]. With its broad use, it is important to understand how consumers share their experience of taking HUMIRA to help gain more insights into its effects directly from users of HUMIRA. Studies have shown that information shared by patients can be different from what was documented by healthcare professionals: a better understanding of adverse experience, better explanation of the nature, significance and consequences, and more detailed information [3].

¹ Corresponding Author, Keyuan Jiang, E-mail: kjiang@pnw.edu.

One potential source of obtaining user experience with medications is social media, including general-purpose platforms such as Twitter. A great number of efforts have been made in using social media as an alternative source of information for pharmacovigilance. As reported by Golder and colleagues in early 2015, more than 3000 published citations were found in their searches of research articles containing “social media” and “adverse events” [4]. Many published works focus on identifying expressions of medication and adverse effect concepts in the social media data, and little effort has been made in understanding the relations between medication and effect, yielding probably incorrect, false-positive or irrelevant results.

A number of semantic relations exist in the medical domain. The U.S. National Library of Medicine (NLM) compiled a collection of normalized semantic relations for biomedical literature², and many of them pertain to medication-effect relations such as *treats*, *causes*, *occurs_in*, *exhibits*, and *produces*. The SemRep software developed by NLM [5] extracts predicates from subject-predicate-object triples found in biomedical citations and maps them to the semantic relations. However, biomedical citations are written with closely followed grammatical and spelling rules, which is not the case for social media posts, in particular Twitter where these rules are not followed by its users to place needed text within the space limitation.

Identifying semantic relations between medication and effect from social media is a challenging task in natural language processing (NLP). Methods developed for formal writing do not perform satisfactorily with social media posts not following grammatical and spelling rules. Mikolov and colleagues reported that using relational similarity in the vector space model of neural embedding of text achieved state-of-art performance in many NLP tasks such as identifying syntactic and semantic relations [6]. To address the limitation, we devised a method of identifying semantic relations by finding the relations most similar to any known semantic relations. In other words, any pair of semantic relations share sufficient similarity is considered semantically similar, indicating that they have similar meanings. In the identified semantic relations between medication and effect, the effects associated with the relations are identified effects.

2. Method

Our relational similarity-based approach infers any potential relations of a study medication based upon their similarity to the known relations of other medications. For example, our method infers a relation of *Depakote: bad dreams* from the known relation of *Toprol XL: chest pain* for that there exists a similarity between the two medication-effect relations. Using this approach, we were able to infer the effects of HUMIRA. Described below are the steps of our pipeline of gathering, processing, analyzing data and validating the results.

2.1. Inference with Relational Similarity

Mikolov and colleagues suggested that in the vector space model, the relational similarity can be derived by computing the cosine value between the two relation vectors each of which is an offset vector of two-term vectors [6]. In our study, to compute the relational similarity, a vector space model was first learned from a corpus of unlabeled

² https://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html

tweets with the help of word2vec software³, and each term, including medication and effect in the text corpus was represented as a real number vector. A medication-effect relation was represented as the offset of the medication vector and effect vector, and the similarity between the two offsets (relation) vectors was computed by the cosine value between the two vectors. Mathematically, we have $med_{base} : effect_{base} :: HUMIRA : effect_{potential}$ to represent the similar relations between two pairs of medication and effect. In implementation, $effect_{potential}$ is derived by finding the most similar effect vector to the result of $vector(effect_{base}) - vector(med_{base}) + vector(HUMIRA)$, where med_{base} and $effect_{base}$ are known and for all the medications except HUMIRA – that is, multiple known relations may be used to infer a single relation. A collection of HUMIRA effects was gathered by inferring all possible effects based upon all the known medication-effect relations of other medications.

2.2. Data

A corpus of 53 million unlabeled tweets related to a list of medication names including HUMIRA was collected by a self-made crawler for twitter.com. The tweets were posted in a time span from March 2006 to summer 2017. The collected tweets were preprocessed to remove any non-English and duplicate tweets. The preprocessed tweets were further filtered by a list of effect terms found in the SIDER database to ensure that each tweet contains HUMIRA and at least an effect expression so that both will appear within the window of the context, facilitating the inference. The resultant 3.6 million “clean” tweets were used to learn the Word2vec vector space model.

SIDER is an online resource of side effect information from marketed pharmaceutical products, and hosted and maintained by EMBL [7]. Its data of all known effects were utilized to filter preprocessed tweets. The SIDER’s effect data of each medication were used to come up with a collection of known medication- effect relations. SIDER has two distinct types of effects: side effects (SEs) and indications (INDs). One major advantage of the SIDER data is that each effect is associated with a UMLS Concept Unique Identifier (CUI), facilitating the expansion of effect concepts to include corresponding Consumer Health Vocabulary (CHV) terms.

2.3. Annotation

Inferred HUMIRA effects are the results of vector algebraic arithmetic, and the vector manipulations do not appear to be straightforward to comprehend in the semantic sense. Annotating related tweets can help us confirm that inferred effects (or relations) were actually mentioned in the Twitter posts. After effects were inferred, we expanded each effect concept by adding its corresponding CHV terms, and tweets containing both HUMIRA and the CHV-expanded effects were extracted from the corpus of “clean” tweets. The extracted tweets were annotated by one of the co-authors to label them as SE or IND based upon the semantics of the tweet text.

2.4. Verification

To uncover if any inferred effects were known or potentially unreported, we, in early 2018, checked the inferred CHV-expanded effects against six reliable sources: SIDER,

³ <https://code.google.com/archive/p/word2vec/>

FDA Dailymed, Fdable.com, MedlinePlus, PubMed, and Google Scholar. SIDER contained both side effects and indications of many marketed products, but it did not contain either for HUMIRA. FDA dailymed contains both side effects and indications. Fdable.com is a Web-search engine of FDA Adverse Event Reporting System (FAERS) data, containing adverse event data. MedlinePlus provides curated consumer health information and prescription and over-the-counter medication information. Both PubMed and Google Scholar helped us to verify if any HUMIRA were mentioned in any scientific publications. Any inferred effects found in any of the six sources were considered known, and those not found were considered potentially unreported.

3. Results and Discussions

Table 1. Statistics of inferred effects.

Count of Effects		Count of SE Effects	Count of IND Effects
Inferred	87	70	17
Annotated	55	44	11
Accuracy	0.632	0.629	0.647

As shown in Table 1, our relational similarity-based method generated 87 inferred effects, among which 70 were SE effects and 17 IND effects. After annotation, 55 effects were confirmed to have relevant tweets, among which 44 were SE effects and 11 IND effects, yielding an overall accuracy of 0.632.

Table 2. Statistics of inferred effects after annotation and verification.

Effects	Count of Inferred	Count of Known	Count of Unreported
SE	44	36	8
IND	11	10	1

After verification against six reliable sources, 46 out of 55 annotated effects were found to be previously reported, showing the consistency on these effects, and 9 of them were found to be potentially unreported, as shown in Table 2. It demonstrates the ability of our method to (1) correctly infer HUMIRA effects which had been previously observed and reported, and (2) infer effects potentially unreported previously. Being able to infer the known effects correctly indicates that our relational similarity-based method can discover existing effects similar to the known base relations. Capable of inferring potentially unreported effects are of special importance for pharmacovigilance in which the safety or suspected effects of pharmaceutical products are being continuously monitored and assessed. It was noticed that there is no entry of HUMIRA or adalimumab in the SIDER database, and this does not affect the inference of HUMIRA effects. Instead, it shows a unique power of our method: effects of any medications can be inferred whether their data exist in SIDER as long as the data for the study medications can be incorporated into the model.

Table 3. Example tweets of potentially unreported HUMIRA effects


Effect	CUI	Tweets
Ache	C0234238	had an overall <u>ache</u> after my humira shot...Thursday my Cellulitis infection came back.
Panic attacks	C0086769	Day 2 <u>panic attacks</u> and chest pain. Think new drug is freaking out my immune system. #IBD # Humira
Swellings	C0013604	Again humira day and got huge swelling around injection site anyone else get this ??? 

Table 3 lists a few example tweets relevant to the HUMIRA effects inferred by our method. Not only do they mention both HUMIRA (the medication) and effects, but they do show the semantic relations between both. Any potentially unreported effects of HUMIRA discovered by our method warrant further assessment or clinical investigation. They may serve as the foundation of developing a hypothesis of suspected effects to help advance pharmaceutical knowledge, medical science, and clinical practice. It is known that misspellings of medication names and side effect concepts are common in Twitter posts [8, 9]. Covering broad linguistic variations of tweet text is one area of improvement for this project, and it may help discover more HUMIRA effects from Twitter posts.

4. Conclusion

In this work, we investigated mining HUMIRA effects mentioned in Twitter posts using our relational similarity-based method, and we were able to identify effects previously known and potentially unreported, demonstrating the power and effectiveness of our relational similarity approach. We believe that our pipeline of gathering, processing, and analyzing data is applicable to identifying the effects of many other medications mentioned in Twitter posts, to gather user-generated information on medication effects.

Acknowledgement: Authors wish to thank anonymous reviewers for their critiques and constructive comments that helped improve this manuscript. This work was supported by the US National Institutes of Health Grant 1R15LM011999-01.

References

- [1] AbbVie, *2018 Annual Report on Form 10-K and 2019 Proxy Statement*, AbbVie, North Chicago, 2019.
- [2] AbbVie, *Important safety information about HUMIRA® (adalimumab)*, <https://www.humira.com/>, 2013.
- [3] L. Härmark, J. Raine, H. Leufkens, I.R. Edwards, U. Moretti, et al., Patient-reported safety information: a renaissance of pharmacovigilance?, *Drug safety* **39(10)** (2016), 883-890.
- [4] S. Golder, G. Norman, Y.K. Loke, Systematic review on the prevalence, frequency and comparative value of adverse events data in social media, *British Journal of Clin Pharmacol* **80** (2015), 878-888.
- [5] T.C. Rindflesch, M. Fiszman, The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text, *Journal of Biomedical Informatics* **36(6)** (2003), 462-477.
- [6] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems* (2013), 3111-3119.
- [7] M. Kuhn, I. Letunic, L.J. Jensen, P. Bork, The SIDER database of drugs and side effects, *Nucleic acids research* **44(D1)** (2015), D1075-D1079.
- [8] K. Jiang, T. Chen, L. Huang, R.A. Calix, G.R. Bernard, A Data-Driven Method of Discovering Misspellings of Medication Names on Twitter, *Studies in health technology and informatics* **247** (2018), 136-140.
- [9] K. Jiang, T. Chen, R.A. Calix, G.R. Bernard, Identifying Consumer Health Terms of Side Effects in Twitter Posts, *Studies in health technology and informatics* **251** (2018), 273-276.

Ontology-Guided Policy Information Extraction for Healthcare Fraud Detection

Theodora S. BRISIMI ^{a,1,2}, Vanessa LOPEZ ^a, Valentina RHO ^a, Marco SBODIO ^a, Gabriele PICCO ^a, Morten KRISTIANSEN ^b, John SEGRAVE-DALY ^b and Conor CULLEN ^b

^a IBM Research, Ireland

^b IBM Watson Health, GHHS, Ireland

Abstract. Financial losses in Medicaid, from Fraud, Waste and Abuse (FWA), in the United States are estimated to be in the tens of billions of dollars each year. This results in escalating costs as well as limiting the funding available to worthy recipients of healthcare. The Centers for Medicare & Medicaid Services mandate thorough auditing, in which policy investigators manually research and interpret the policy to validate the integrity of claims submitted by providers for reimbursement, a very time-consuming process. We propose a system that aims to interpret unstructured policy text to semi-automatically audit provider claims. Guided by a domain ontology, our system extracts entities and relations to build benefit rules that can be executed on top of claims to identify improper payments, and often in turn payment policy or claims adjudication system vulnerabilities. We validate the automatic knowledge extraction from policies based on ground truth created by domain experts. Lastly, we discuss how the system can co-reason with human investigators in order to increase thoroughness and consistency in the review of claims and policy, to identify providers that systematically violate policies and to help in prioritising investigations.

Keywords. Healthcare fraud, ontology-based information extraction, claims auditing

1. Introduction

Improper payments for Medicaid in fiscal year 2018 were estimated at 9.8% of Medicaid spending (\$36.2 billion) [1]. Examples of improper payments include billing for services exceeding permitted unit limitations, billing separately for services already included in a global fee (unbundling) and billing for medically unnecessary services. Regulations around compliance and accurate billing are described in federal and state policies. Armed with a deep understanding of these policies, FWA investigation units aim at identifying violations in claims submitted for reimbursement by medical providers. However, the sheer volume of claims, benefits, and policy to review, combined with the limited investigation resources and varying skill sets of

¹Corresponding Author: Theodora S. Brisimi, E-mail: theodora.brisimi@ibm.com

²We would like to acknowledge Pierpaolo Tommasi, Carlos Alzate, Spyros Kotoulas, Fabrizio Cucci, Tim Cooper, Mark Gillespie, Mark Gilroy, Jonathon Ryan, Grace Ferguson, John Davis and Jillian Scalvini for their support and insights for this work.

investigators, lead to poor coverage and understanding of the opportunity landscape. The Centers for Medicare & Medicaid Services have developed a fraud detection system that analyses claims to identify providers with suspicious billing patterns [2]. This and similar approaches rely on a combination of applying data analytics techniques to patterns in structured claims data [3, 4] and hand-crafting rules that flag sets of claims for additional investigation. The system presented in this paper focuses on fusing policy and claims together to identify improper payments. It achieves this by semi-automatically converting policy text into *benefit rules*, which are applied directly to claims. This approach is novel to the best of our knowledge and presents a number of research challenges, including:

- a) *Domain modelling & knowledge extraction*: Key to our system is building an ontology that is flexible enough to capture the diversity and complexity of the policy *benefit rules* as well as expressive enough to represent the knowledge in a simple, unambiguous and human-readable way to support policy comprehension and human oversight. The ontology is used to integrate relevant terminology and heterogeneous domain sources (e.g., state programs, eligible places of service, coding systems, such as the International Classification of Diseases ICD-9/10 and corresponding updates), as well as to guide the information extraction.
- b) *Linking of policy information to the claims data*: Some elements of the *benefit rules* can be easily mapped to claims data columns (e.g., a minimum age constraint will be linked to the birth date of the patient), while others require a more complicated process (e.g., aggregated units of service for a given patient over a given period). The system also needs to map services described in policy to their representation form used in claims, such as the Healthcare Common Procedure Coding System (HCPCS).
- c) *Rules polarity*: In policy documents, benefit rules usually express what will be reimbursed (a *positive* rule), e.g. "Adults at high risk for caries may have up to 2 units of dental prophylaxis per year". Sometimes, they express what will not be reimbursed (a *negative* rule), e.g. "Dental prophylaxis is not a covered benefit for children aged 0-4". This provides a challenge for automated identification of improper claims because while the negative rules are generally unambiguous and directly identify violations, the positive rules are often ambiguous and incomplete. In the example, we must infer that more than two units are not allowed for a high-risk adult and that there may be an implied rule for non-high-risk adults and separate rules for children and high-risk children.
- d) *Validation of the approach*: Validating the feasibility and impact of our system when used as part of the investigators' workflow requires constructing a set of ground truth rules to evaluate its performance and measuring its generalizability across different policy areas as well as geographic regions.

2. Method

Below we present the components that comprise our system. Fig. 1 runs an example through the end-to-end pipeline.

Policy Ingestion: This component converts the PDF policy document into an HTML document ³, and further enriches the HTML with information identifying sentences, and their hierarchical arrangement (e.g., grouping sentences under a common header or identifying a paragraph introducing a list).

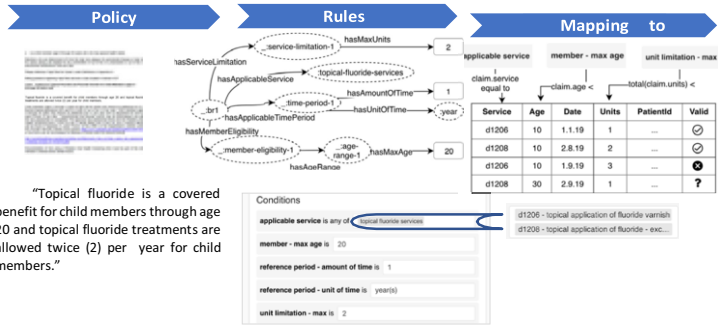


Figure 1. Proposed system's pipeline overview with *benefit rule* extraction and claims validation example.

Rules Extraction: This component is an ontology-based information extraction system based on the enriched HTML policy. Two extractors are implemented on top of two different NLP technologies: Watson X [5] and SystemT [6]. For each textual fragment, the extraction involves the annotation of medical concepts and relations guided by UMLS and the domain ontology. We construct a machine-processable representation of a benefit rule by reasoning over the ontology and translating textual patterns into ontological entities and corresponding relations. Each extracted rule is represented in a knowledge graph (KG). The rule extraction process may generate multiple KGs from the same textual fragment(s), which can be consolidated based on different strategies. Human oversight of this extracted knowledge is essential to correct any errors and establish trust. To that end, we transform each KG into a user-friendly representation, that contains conditions and corresponding values (e.g., *applicable service* with value *topical fluoride* or *maximum age* with value 20). This empowers the investigators to interact with, adapt and use the extracted knowledge. Rules in policies follow different templates that we captured in the ontology working together with FWA investigators. A detailed discussion of the ontology creation is presented in [7]. The extracted *benefit rules* are then normalised, i.e., properties values are standardised, so that rules execution can unambiguously identify them in claims. The normalisation steps include linking extracted services to specific procedure codes, e.g., *topical fluoride treatments* maps to codes *D1206* and *D1208*, and representing dates in a standard format, e.g., representing *fiscal year* as the appropriate dates range depending on the geographic area.

Mapping to claims: The conditions of each *benefit rule* must be mapped to claims columns in order for the rule to be executed. We assume a claims schema against which the mapping is performed. Examples of mappings are: direct - e.g., a minimum age property in the rule maps directly to an age column that can be calculated from the birth date of the patient and the date of service, aggregations - e.g., a rule that places an upper limit on the number of claims over a period of time. To map this to claims, we need to (i) filter claims, e.g. by service, (ii) group claims, e.g., by patient or date and (iii) compute the aggregation, e.g. an incremental count of claims, value-dependent - e.g., a rule may have a 'requirement' element with values such as 'prior authorisation', 'medical necessity', etc. Each of these values may map to distinct columns in the claim data.

³IBM Compare and Comply: www.ibm.com/watson/developercloud/compare-and-comply/

This component applies the *benefit rules* to claims and outputs a status for each claim: improper, correct or unknown, with the latter representing claims that are not represented by any rule. Claims that appear to be improper are linked automatically to both the executed benefit rule and the relevant policy documentation.

3. Results and Discussion

We designed a prototype investigation workbench (Fig. 2) that highlights the impact of bringing policy and claims data together into a unified workflow. Using this workbench, investigators can contrast what providers are authorised to do with what they actually do. This starts by selecting a policy area (e.g., topical fluoride) and the policy sections (rules) to be investigated (Fig. 2: #1, #2). An Investigator can then immediately visualise, inspect and prioritise claims that appear to be out-of-sync with that policy (Fig. 2: #3,#4).

This helps in addressing significant problems in investigation planning and execution: (1) investigators can size potential recovery opportunities before deciding where to allocate scarce resources; (2) maintaining a direct connection between policy and related claims helps build a water-tight case for recovery; (3) automatically showing the relevant claims (and claim fields) for a policy area under investigation saves time by reducing dependencies on other departments for pre-defined data extracts/spreadsheets; and (4) it enables investigators to review and correct the automatically-extracted benefit rules (Fig. 2: #2), thereby building up a shared store of high-quality, executable policyknowledge.

We validate the rules extraction performance via precision and recall. To create Ground Truth (GT) for these measurements, we used a set of documents for Physical Therapy policy from one U.S. state, plus Dental policy from two U.S. states. For each one, a team of three FWA Investigators manually translated the associated benefit rules into a computable form, based on our ontology (**GT rules**). Automatically **extracted rules** were then compared to these, yielding the evaluation results presented in Table 1.

Table 1. We pair a GT rule R with the set of extracted rules from the same text, and we compute a pairing score that takes into account the number of common conditions and values. If the pairing score is always zero, then R is a false negative. Otherwise, the true positive corresponding to R is the extracted rule having the maximum pairing score. Details on the pairing score calculation can be found in [7]. The last column of the table shows the average maximum pairing score across all rules.

Policy	num. GT rules	num. extracted rules	precision	recall	avg. pairing score
Physical Therapy	25	38	0.58	0.88	0.61
Dental (State 1)	50	46	0.65	0.60	0.69
Dental (State 2)	34	42	0.64	0.79	0.56

The ontology used in the experiments comprises of 35 classes, 1151 individuals, 29 properties and 4214 lexicalisations (i.e., ontology entity labels used to annotate textual entities). We are currently extending our work to improve the breadth and coverage of the models and the extraction, in particular across paragraphs, headings and in cases where there are conflicting extracted information. Most of the effort required to generalise across domains and geographical areas is on identifying external (instance) area-specific data to be incorporated into the ontology, such as programs and grouping of codes that are not

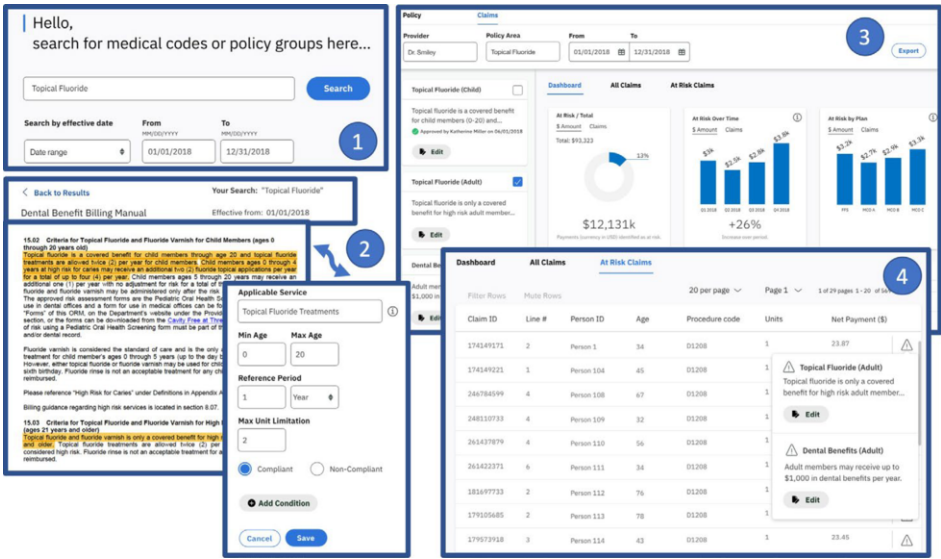


Figure 2. Workflow: The investigators receive a phone tip about a suspicious dental provider *Dr. Smiley* and decide to look for violations related to topical fluoride. Next, they 1) search for related policies, 2) select one or more extracted rules from these policies and modify/correct the automatically extracted rule(s), if necessary, 3) define the scope (e.g., provider and date range), and apply the set of rules to the related claims, seeing the impact of the violations in terms of number of improper claims and potentially recoverable amount of money, 4) carefully inspect each improper claim and the corresponding policy it appears to violate.

part of a global (e.g., federal) code set. Without these groupings, for example, considering the *Dental (State 2)* policy, recall started at around 0.59. When clinical vocabularies such as UMLS are used to find services, treatments and diagnoses, recall increases to 0.76 (with a precision of 0.68). Furthermore, when 135 different groupings of codes were added by a coding expert, recall improved to 0.79, with a small drop in precision to 0.64 (Table 1).

4. References

[1] U.S. Government Accountability Office, https://www.gao.gov/key_issues/medicaid_financing_access_integrity, 2018.

[2] U.S. Government Accountability Office, <https://www.gao.gov/assets/690/686849.pdf>, 2017.

[3] H. Joudaki, et al., Using data mining to detect health care fraud and abuse: a review of literature, *Glob. J. Health Sci.* **7**(1) (2015), 194.

[4] D. Thornton, et al., Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detection, *Proc. Technol.* **9** (2013), 1252–1264.

[5] A. Kalyanpur, et al., Structured data and inference in DeepQA, *IBM J. Res. Dev.* **56.3.4** (2012), 10–1.

[6] L. Chiticariu, et al., SystemT: Declarative text understanding for enterprise, *NACCL-HLT.* **3** (2018), 76–83.

[7] V. Lopez, et al., *Benefit graph extraction from healthcare policies*, ISWC, 2019.

Text-Mining Services of the Swiss Variant Interpretation Platform for Oncology

Déborah CAUCHETEUR^{a,b,*}, Julien GOBEILL^{a,b,*}, Anaïs MOTTAZ^{a,b,*}, Emilie PASCHE^{a,b,*}, Pierre-André MICHEL^{a,b}, Luc MOTTIN^{a,b}, Daniel J. STEKHOVEN^{b,c}, Valérie BARBIÉ^b and Patrick RUCH^{a,b}
^a HES-SO / HEG Geneva, Information Sciences, Geneva, Switzerland
^b SIB Swiss Institute of Bioinformatics, Geneva, Switzerland
^c NEXUS Personalized Health Technologies, ETHZ, Zürich, Switzerland

Abstract. The Swiss Variant Interpretation Platform for Oncology is a centralized, joint and curated database for clinical somatic variants piloted by a board of Swiss healthcare institutions and operated by the SIB Swiss Institute of Bioinformatics. To support this effort, SIB Text Mining designed a set of text analytics services. This report focuses on three of those services. First, the automatic annotations of the literature with a set of terminologies have been performed, resulting in a large annotated version of MEDLINE and PMC. Second, a generator of variant synonyms for single nucleotide variants has been developed using publicly available data resources, as well as patterns of non-standard formats, often found in the literature. Third, a literature ranking service enables to retrieve a ranked set of MEDLINE abstracts given a variant and optionally a diagnosis. The annotation of MEDLINE and PMC resulted in a total of respectively 785,181,199 and 1,156,060,212 annotations, which means an average of 26 and 425 annotations per abstract and full-text article. The generator of variant synonyms enables to retrieve up to 42 synonyms for a variant. The literature ranking service reaches a precision (P10) of 63%, which means that almost two-thirds of the top-10 returned abstracts are judged relevant. Further services will be implemented to complete this set of services, such as a service to retrieve relevant clinical trials for a patient and a literature ranking service for full-text articles.

Keywords. Precision medicine, literature, variant, terminology, text-mining

1. Introduction

After several years of collaboration to improve and harmonize the NGS (next-generation sequencing) practices in somatic mutation calling, a number of Swiss hospitals, pathology institutes and the Swiss Institute of Bioinformatics members have pointed to a set of shortcomings, most prominently the lack of a central repository for clinically verified variant annotations in cancer. To support the harmonization of variant annotation in diagnosis, to provide a centralized curated database of somatic variants coming from Swiss hospitals, and to enable a seamless interaction between participating institutes and global initiatives, the Swiss Variant Interpretation Platform for Oncology (SVIP-O) was developed [1].

¹ Corresponding Author: Deborah Caucheteur.; E-mail: deborah.caucheteur@hesge.ch.

* contributed equally

The curation of variants is a labour-intensive process. First, NGS analyzes result in a huge amount of variants in the patient sample. Second, for each variant, many sources must be consulted, including the scientific literature. While for some already well-studied variants (e.g. BRAF V600E), identifying a set of relevant scientific literature is not an issue, for variants of uncertain significance (VUS), the task can be more difficult. Indeed, the curator must comprehensively gather the relevant literature to assign a standardized tier level (ASCO/AMP/CAP guidelines), in which the variant, the gene or the diagnosis could have been labelled in different ways [2]. Thus, the curator must author multiple queries to avoid missing out on an important paper. Moreover, when large sets of literature are available, triage of the literature (i.e. selection of relevant papers as well as rejection of irrelevant papers) can be a very time-consuming task [3].

Therefore, we developed a set of text analytics services intended to facilitate and improve the comprehensive collection of literature to support further processing steps, which will include the capture of textual evidence by Swiss-Prot curators and clinical validation by medical specialists for final storage in the SVIP-O knowledgebase. First, we developed a service to annotate the clinically relevant information contained in the scientific literature. This service enables not only to accelerate searches (i.e. queries are performed using unique identifiers in a standardized field) but also to increase the recall (i.e. the annotation step conflates synonyms and string variants). However, the annotation of variants faces a serious limitation: no universal terminology is available. In order to offset this issue, variants can be searched in literature using synonym expansion. Thus, as a second task, we propose a service to expand a variant name with a set of expressions, including Human Genome Variation Society (HGVS) standard descriptions for the protein, transcript and genomic DNA levels, but also non-standard formats found in the literature (e.g. V-600-E or BRAFV600E) [4]. Finally, a service collecting and prioritizing literature is available: it combines heterogeneous information retrieval results for an optimal article triage, which altogether can reduce the triage effort by a factor of 3 [5].

2. Methods

Our MEDLINE and PMC collections consisted of respectively 30,415,832 and 2,632,396 documents (in January, 7th 2020), daily updated and loaded into a MongoDB document database and then into an ElasticSearch index.

2.1. Recognition of terminological entities in the literature

The collection was annotated with codes from various terminologies and ontologies, such as neXtProt [6] for genes, Drugbank [7] and WHO-ATC [8] for drugs, NCI Thesaurus [9] and ICD-O-3 [10] for diseases, and HPO (Human Phenotype Ontology) for phenotypes. Querying MEDLINE through annotations is not only faster because the indexes are pre-computed but also it results in a better recall as each occurrence of a concept receives one unique ID for all its synonyms. We also applied string pre-processing methods. For instance, if dashes are present, they are treated to form a “new” word (e.g. “AB-C” becomes “AB”, “C” and “ABC”). Papers which contain only the word without dash will thus be retrieved.

2.2. Generation of synonyms for variants

While many databases of polymorphisms and variants exist, such as ClinVar, COSMIC or dbSNP, using those resources as terminologies is fairly challenging. They describe variants using a standard nomenclature recommended by the HGVS [11]. These standards require a precise syntax and a reference sequence on which the variation is described to avoid uncertainty about the position of the change. Both of them are rarely observed in publications. Depending on the database, variants entries are also centered on different levels: genomic, transcript or protein, which are not true synonyms.

Therefore, we developed our own synonym generation tool that enables us in addition to annotate variants not necessarily present in those databases. Our efforts focused until now on SNPs. Our tool generates synonyms given a gene and a variant. It includes a validation step, where we check whether the given base or amino acid exists at the given position. Then, we compute the description of the variant at the other levels, using the Mutalyzer tool [12]. We finally generate synonyms with many syntactic variations as encountered in the literature [4]. We also extend the search to all possible mutations in the case the replacing amino acid is not defined.

2.3. Prioritization of literature

The report focuses on the literature triage tasks, i.e. the ability to rank MEDLINE abstracts - as opposed to full-text articles - to support further curation steps. Triage is usually performed on abstracts, whose content is sufficient to help a domain expert to decide whether a particular report needs in-depth reading or not.

Our literature prioritization system is based on two steps: collecting a complete set of abstracts and reranking the MEDLINE set. A most complete set of abstracts related to a particular triplet (i.e. a variant in a gene for a specific diagnosis) is built by the intersection of several queries' output: normalized entities (i.e. gene and diagnosis) are searched with unique identifiers within the MEDLINE annotations, while a keyword search expanded with synonyms is performed within free texts from MEDLINE for not normalized entities (i.e. variant). Moreover, a set of queries, with decreasing levels of specificity (e.g. abstracts not mentioning diagnosis) is also performed. Results are linearly combined with previous abstracts set. Then, we apply different strategies to re-rank the MEDLINE set: 1) based on the number of occurrences of some annotated entities (e.g. abstracts mentioning drugs); 2) based on demographic information; 3) based on a set of keywords that should (e.g. treat) or should not (e.g. marker) be present in the abstracts and 4) based on the breadth of treatments returned in the top articles. The last re-ranking strategy aims at avoiding that all top returned articles are related to the same treatment, but rather favoring abstracts that are discussing different treatment options.

The system is evaluated following standard TREC procedures using TREC PM 2019 benchmarks [13], comprising 40 synthetic patient cases, each consisting of a disease, a variant, a gene and some demographic information.

3. Results

3.1. Recognition of terminological entities in the literature

Thanks to our annotations, we are able to retrieve more papers than classical queries on PubMed. The query *BRAF* retrieves 14,099 papers on PubMed versus 14,952 papers on our index with his corresponding code *NX_P15056*. Another example, query *non-small cell lung cancer* retrieves 51,858 papers on PubMed versus 70,972 (code *C2926*) with our index. These results are explained by expansions provided with annotations: synonyms (*BRAF1* or *RAFB1*; *NSCLC*) and processing of hyphens (*B-RAF* becomes *BRAF*, thus a paper containing *B-RAF* will be retrieved). Table 1 represents an overview of annotations which are available for our collections for some of the terminologies.

Table 1. Extract of statistics about annotations on MEDLINE/PMC collections obtained in January 2020.

Type	Terminologies	No of entities annotated in	
		MEDLINE	PMC
Drugs	DrugBank	80,100,2684	86,769,820
Drugs	ATC	46,173,559	30,404,688
Diseases	NCIt	131,577,959	108,049,190
Diseases	ICD-O-3	4,837,713	2,955,200
Genes	neXtProt	36,320,459	91,930,787
Total annotations		785,181,199	1,156,060,212
Average per document		26	425

3.2. Generation of synonyms for variants

Our tool generates 42 synonyms for a given valid variant when successfully mapped to all levels, with 16 synonyms for protein variant, 13 for transcript variant including COSMIC id and 13 for genomic variant including dbSNP id. It increases the retrieval between a few percent and several times the number of publications depending on the variant frequency, with smaller effect for very popular variants. For instance, 708 additional abstracts, over 1715, are retrieved for the variant V617F in JAK2.

3.3. Prioritization of literature

Precision at rank 10 (P10) has been used to evaluate our system. This metric reflects the proportion of relevant documents retrieved in the top ten results. Our system resulted in a P10 of 63%, which means that almost two thirds of the top-10 returned abstracts are judged relevant. This service is publicly available: <http://candy.hesge.ch/Variomes/>.

4. Discussion

We have thus developed a set of services that can be used to facilitate the process of variant curation and validation by physicians, and in particular molecular pathologists, oncologists and hematologists, as well as rare disease experts. The literature triage service, boosted by the variant expansion service and the MEDLINE annotations service, is able to reduce the search burden by simplifying the paper triage. Indeed, the system's

evaluation demonstrated that in the top-10 abstracts proposed by our system, more than six are relevant for the clinical decision-support task. In addition, such services can increase the possibility of finding a relevant paper. Although this aspect might be marginal for common variants, it is key for rare (or poorly studied) variants for which each publication matters. As an example, the search for the TP53 V143A variant in PubMed results in only seven abstracts, while our system is able to return 34 abstracts. Indeed, while PubMed strictly searches for V143A, our system is also able to search for Val143Ala or 428T>C.

Additional services will be implemented to complete this set of services, such as a literature ranking service for full-text articles. Indeed, scientific abstracts reporting on treatments do not always mention all the information regarding diagnosis, gene and variant, so that full-text articles are needed, as well as supplementary data when available. A service to retrieve relevant clinical trials for a given patient is also under development, enabling to connect patients with experimental treatments.

Acknowledgement: This project has been supported by Swiss Personalized Health Network (SPHN) and BioMedIT fundings, see <https://svip.ch/>.

References

- [1] D.J. Stekhoven, P. Ruch, V. Barbié, Swiss Variant Interpretation Platform for Oncology (SVIP-O), *Swiss Med Informatics* **34** (2018), 00411.
- [2] Z. Shi, B. Gu, F. Popowich, et al., Synonym-based Query Expansion and Boosting-based Re-ranking: A Two-phase Approach for Genomic Information Retrieval, In *Proceedings of the Fourteenth Text REtrieval Conference TREC* (2005).
- [3] T.C. Wieggers, A.P. Davis, K.B. Cohen, et al., Text mining and manual curation of chemical-gene-disease networks for the Comparative Toxicogenomics Database (CTD), *BMC Bioinformatics* **10** (2009), 326.
- [4] Y.L. Yip, N. Lachenal, V. Pillet, et al., Retrieving mutation-specific information for human proteins in UniProt/Swiss-Prot Knowledgebase, *J Bioinform Comput Biol* **5(6)** (2007), 1215-31.
- [5] L. Mottin, J. Gobeill, E. Pasche, et al., neXtA5: accelerating annotation of articles via automated approaches in neXtProt, *Database (Oxford)* (2016).
- [6] P. Gaudet, P.A. Michel, M. Zahn-Zabal, et al., The neXtProt knowledgebase on human proteins: 2017 update, *Nucleic Acids Res* **45(D1)** (2017), D177-D182.
- [7] D.S. Wishart, Y.D. Feunang, A.C. Guo, et al., DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res* **46(D1)** (2018), D1074-D1082.
- [8] WHO Collaborating Centre for Drug Statistics Methodology, ATC classification index with DDDs, 2019. Oslo, Norway (2018).
- [9] N. Sioutos, S. de Coronado, H.W. Haber, et al., NCI Thesaurus: a semantic model integrating cancer related clinical and molecular information, *J Biomed Inform* **40(1)** (2007), 30-43.
- [10] A. Fritzl, C. Percy, A. Jack, et al., International classification of diseases for oncology / editors, 3rd ed. *World Health Organization* (2000).
- [11] J.T. den Dunnen, R. Dalgleish, D.R. Maglott, et al., HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat* **37(6)** (2016), 564-9.
- [12] M. Wildeman, E. van Ophuizen, J.T. den Dunnen, et al., Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker, *Hum Mutat* **29(1)** (2008), 6-13.
- [13] K. Roberts, D. Demner-Fushman, E.M. Vorrhees, et al., Overview of the TREC 2018 Precision Medicine Track, In *Proceedings of the Twenty-Seventh Text REtrieval Conference* (2018).

The Smart Device System for Movement Disorders: Preliminary Evaluation of Diagnostic Accuracy in a Prospective Study

Julian VARGHESE ^{a,1}, Michael FUJARSKI ^a, Tim HAHN ^b, Martin DUGAS ^a, Tobias WARNECKE ^c

^a Institute of Medical Informatics, University of Münster, Germany

^b Department of Psychiatry, University Hospital of Münster, Germany

^c Department of Neurology, University Hospital of Münster, Germany

Abstract. Consumer wearables can provide objective monitoring of movement disorders and may identify new phenotypical biomarkers. We present a novel smartwatch-based prototype, which is implemented as a prospective study in neurology. A full-stack Machine Learning pipeline utilizing Artificial Neural Networks (ANN), Random Forests and Support Vector Machines (SVM) was established and optimized to train for two clinically relevant classification tasks: First, to distinguish neurodegenerative movement disorders such as Parkinson's Disease (PD) or Essential Tremor from healthy subjects. Second, to distinguish specifically PD from other movement disorders or healthy subjects. The system was trained by 318 samples, including 192 PD, 75 other movement disorders and 51 healthy participants. All models were trained and tested with hyperparameter optimization and nested cross-validation. Regarding the more general first task, the ANN succeeded best with precision of 0.94 (SD 0.03) and recall of 0.92 (SD 0.04). Concerning the more specific second task, the SVM performed best with precision of 0.81 (SD 0.01) and recall of 0.89 (SD 0.04). These preliminary results are promising as compared to the literature-reported diagnostic accuracy of neurologists. In addition, a new data foundation with highly structured and clinically annotated acceleration data was established, which enables future biomarker analyses utilizing consumer devices in movement disorders.

Keywords. Mobile applications, artificial intelligence, machine learning, movement disorders

1. Introduction

Movement Disorders as Parkinson's Disease (PD) are primarily diagnosed via clinical examination. A meta-analysis by Rizzo et al. 2016 indicated that the pooled diagnostic accuracy is at 73.8% for general neurologists, geriatricians, or general practitioners (95% Credible Interval (CRI): 67.8%-79.6%) and 79.6% (95% CrI 46%-95.1%) for movement disorder experts regarding initial assessment and 83.9% (95% CrI 69.7%-92.6%)

¹ Corresponding Author, Julian Varghese, Institute of Medical Informatics, University of Münster; E-mail: julian.varghese@uni-muenster.de

regarding refined diagnosis after follow-up [1]. It ascertains that the overall validity of clinical diagnosis did not improve in the last 25 years and new biomarkers are needed, ideally with high availability. We present a novel mobile prototype, called the smart device system (SDS), which has been implemented and approved for a two-year prospective study [2]. The system is the first one, which captures two-side synchronous acceleration data from smartwatches with the integration of electronic questionnaires to account for early symptoms, medication, medical and family history. Health informaticians and movement disorder specialists designed the system, including a smartphone app to guide the examination step by step. Previously, we have demonstrated the accuracy of raw measurement data and tremor frequency detection as proof of concept to visualize and detect subtle movement characteristics, which are imperceptible to human vision [3].

As a clinical use case, we trained a diagnosis classification model. To our knowledge, most of the related models (summarized in [4]) for PD deal with limited sample sizes ($n < 100$) and are overly restricted to classify PD vs healthy. However, in clinical practice, the neurologist does not only face people being exclusively a PD case or healthy case. Instead, there are many differential diagnoses with PD-similar symptoms. In these cases, the models mentioned above would not be applicable or would perform significantly poorer. To address this issue, an important disease category is part of the patient recruitment: movement disorders other than PD as essential tremor, atypical parkinsonism, secondary causes of parkinsonism, dystonia, which will be summarized as DD in the following. Our system is trained for two clinically relevant classification tasks. First, to distinguish between movement disorders (PD+DD) and healthy subjects. Such a system could potentially be used at home-based settings or at general practices, e.g. to indicate whether certain abnormal movement characteristics (e.g. hand tremor) is pathologic or still normal (e.g. physiological tremor). Second, to distinguish PD from the rest (DD + healthy), which addresses a setting in specialized practices.

2. Methods

Details of the study design, examination steps and feature extraction have been published previously [2]. Study participant's assessment starts with smartphone-based data entry on medication, current diagnosis, family history and early symptoms. The smartwatch-based examination consists of ten different coordination tasks, including serials seven to provoke and monitor different movement characteristics while seated in an armchair. Altogether, the examination takes 15 minutes. The system is composed of two smartwatches (Apple Watch Series 4) for two-hand acceleration measurement and two smartphones (Apple iPhone version 7) for electronic questionnaires, examination guidance, watch-pairing and data transfer. These devices are connected via one central iOS app on the master smartphone. Sampling rate of acceleration data in all three axes is 100 Hz. Each sample is pseudonymized and ready for real-time visualization and further data processing [3]. Table 1 summarizes the three aforementioned relevant study participants categories PD, DD and Healthy. The full list with demographic details and years of disease onset is provided for each sample in the supplement. It also includes diagnosis details of the DD cluster [5]. All diagnoses were confirmed by neurologists and finally reviewed by one senior movement disorder expert. The current stack of the data analytics pipeline is implemented in Python, illustrated in Figure 1 and includes

Artificial Neural Networks (ANN), Random Forests and Support Vector Machines (SVM) as the three main classifiers.

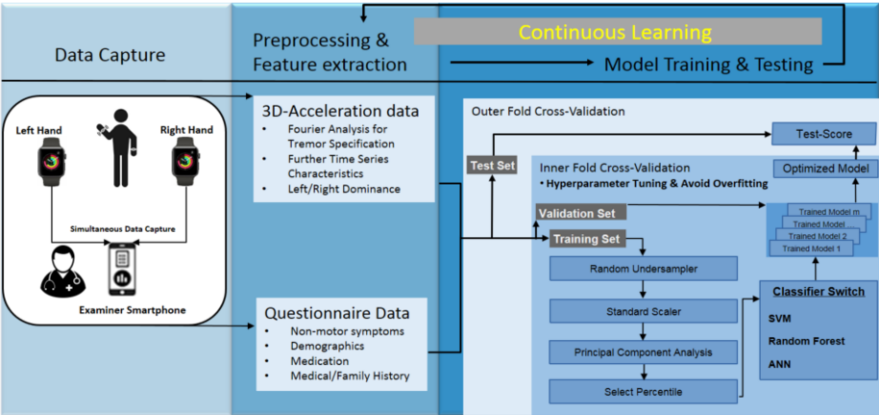


Figure 1. Data Analytics pipeline. ANN = Artificial Neural Networks, SVM= Support Vector Machine.

Table 1. Class population.

Class	Sample Size	Average Age (SD)
PD	192	65.71 (9.61)
DD	75	60.88 (15.58)
Healthy	51	60.45 (14.22)

TSFresh 0.12.0 is used to extract a series of general-purpose time-series characteristics [5]. Moreover, acceleration data is processed with Fast Fourier Transform to extract tremor frequencies. Amplitude-Histogram analyses are applied and checked for dominances on the left or right-hand side. A complete list of all extracted features and descriptions is given in the supplement [6]. The set of features is trained with nested cross-validation using five inner and five outer folds. During one inner fold, we applied the random undersampler from Scikit Learn 0.21.3 [7] in order to remove the bias towards the majority class by randomly removing samples of that set. The standard scaler from Scikit Learn subtracts the mean and scales to the unit variance for every feature. The principal component analysis (PCA) reduces the dimensionality without assuming independence. The Scikit Learn-based 'Select Percentile' step randomly selects a subset of features, which are then used for training the classifier. While the standard Scikit Learn library implements SVM and Random Forests, the ANN is implemented with dense multi-layered perceptron networks using the KERAS 2.2.4 package and Google's Tensorflow 1.3.1, which provides full GPU support [8]. The whole process of model training and testing is wrapped by Photon-AI, a hyperparameter optimization toolbox [9]. During an inner fold, we optimize the hyperparameters for the PCA, the Select Percentile and the classifiers. We used the Scikit Learn's grid-search optimizer as our optimizer. A list of the full hyperparameter space is provided in our supplement [6].

3. Results

Tables 2 and 3 list the best performing models for each of the three classifiers after nested cross-validation. While the ANN performed best for the first task regarding average accuracy of 0.89 and precision of 0.94, the SVM – using radial basis function (RBF) as kernel – performed best in the second task with accuracy of 0.79 and precision of 0.81. None of the different classifiers outperformed the others when taking the overlap of their corresponding standard deviations into account. Details on all optimized hyperparameter values and best performing features are provided in the supplement [6]. Further classifiers, including deep neural networks, are being evaluated and discussed in the next section.

Table 2. Average Performances for general classification task: Parkinson's and related movement disorders vs Healthy. SD= standard deviation. Rbf= Radial Basis function.

Estimator	Accuracy (SD)	Precision (SD)	Recall (SD)	F1 (SD)
ANN	0.89 (0.00)	0.94 (0.03)	0.92 (0.04)	0.93 (0.00)
Random Forest	0.88 (0.04)	0.91 (0.04)	0.95 (0.04)	0.93 (0.02)
SVM - rbf	0.89 (0.03)	0.92 (0.03)	0.95 (0.04)	0.93 (0.02)

Table 3. Average Performances for specific classification task: PD vs. Other related movement disorders and Healthy.

Estimator	Accuracy (SD)	Precision (SD)	Recall (SD)	F1 (SD)
ANN	0.77 (0.04)	0.78 (0.04)	0.90 (0.06)	0.83 (0.03)
Random Forest	0.75 (0.03)	0.75 (0.03)	0.91 (0.04)	0.823 (0.02)
SVM - rbf	0.79 (0.01)	0.81 (0.01)	0.89 (0.04)	0.85 (0.01)

4. Discussion

The more general task was evaluated with precision and recall above 90% for all classifiers, which highlights potential applicability in non-specialized settings as home-based assessment or general practices. Regarding the more specific task, the average diagnostic accuracy (79%) of the best performing model is higher than pooled reported accuracy of general practitioners and comparable to movement disorder experts regarding initial assessment [1]. A number of machine-learning approaches share a major disadvantage as they are trained by retrospective data sources, with high data quality, which however might be illusory at real-time or point of care. A major strength of our mobile system is that it only requires the self-generated data from the 15 minutes examination being not dependent on any further patient data or external information systems. Thus, we are confident that our results are reproducible in prospective scenarios. The reported results still need to be taken with caution as our setting comes with general limitations: The study is conducted at a single site, and thus we cannot rule out bias due to specific site population, examination setting or target diagnoses. We recruit from a large tertiary care centre and include all types and stages of Parkinson's disease and other movement disorders, and therefore, we believe the training data provides a suitable data foundation to start from. However, as a matter of our recruitment

setting, only a minority of samples are Parkinson's cases with early disease stage, which would be the ideal population to train from. After study completion, a multi-centric study design is planned for 2021, which would include different outpatient clinics and home-based assessments. At the current state, the system is to be used for scientific purpose only as it is not cleared as a medical device.

We have tested several approaches towards Deep Learning with recurrent networks as Long-Short-Term Memories. Their performances were remarkably poorer, being not significantly higher than random guessing. Our sample size may not be sufficient (e.g. > 10,000) to train a deep network that requires substantially more model parameters. As our data-preprocessing steps evolve and the sample size grows, we are monitoring different types of Neural Networks and further traditional statistical models. Apart from the machine-learning framework, the highly interactive research platform enables researchers to visualize objective movement characteristics and to have a deeper look at subtle characteristics within a hundredth of a second. In a previous study, we could visualize tremor characteristics imperceptible to non-sensor-based human vision [3]. This could unveil further morphological features in the future, which are not yet utilized.

To conclude, we have generated a sensor-based data foundation to study movement characteristics in a broad area of different movement disorders. Machine Learning classifiers show high accuracy for distinguishing relevant diagnoses classes compared to expert-based performances. Further research is underway to show generalizability of the system.

5. Acknowledgement

This work is funded by the Innovative Medical Research Fund (Innovative Medizinische Forschung, I-VA111809) of the University of Münster.

References

- [1] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, G. Logroscino, Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis, *Neurology* **86**(6) (2016), 566–76.
- [2] J. Varghese, S. Niewöhner, I. Soto-Rey, S. Schipmann-Miletić, N. Warneke, et al., A Smart Device System to Identify New Phenotypical Characteristics in Movement Disorders, *Front Neurol* **10** (2019), 48.
- [3] J. Varghese, S. Niewöhner, M. Fujarski, I. Soto-Rey, A.L. Schwake, et al., Smartwatch-based Examination of Movement Disorders: Early Implementation and Measurement Accuracy, *EGMS* (2019). DOI:10.3205/19GMDS136.
- [4] AU. Haq, J.P. Li, M.H. Memon, J. Khan, A. Malik, et al., Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings, *IEEE Access* **7** (2019), 37718–34.
- [5] M. Christ, N. Braun, J. Neuffer, A.W. Kempa-Liehr, Time series feature extraction on basis of scalable hypothesis tests (tsfresh - a python package), *Neurocomputing* **307** (2018), 72–7.
- [6] J. Varghese, *Online Supplement SDS MIE 2020*, <https://uni-muenster.sciebo.de/s/dYv1d7weDar7xnZ>, (accessed 09/31/2019).
- [7] G. Hackeling, *Mastering Machine Learning with scikit-learn*, Packt Publishing, 2017.
- [8] *Keras Documentation*, <https://keras.io>, (accessed 10/01/2019).
- [9] *Photon-AI. A Python-based Hyperparameter Optimization Toolbox*, <https://www.photon-ai.com/>, (accessed 30/09/2019).

User Expectations and Willingness to Share Self-Collected Health Data

Ashenafi Zebene WOLDAREGAY ^{a,*}, André HENRIKSEN ^{a,1,*}, David-Zacharie ISSOM ^b, Gerit PFUHL ^a, Keiichi SATO ^c, Aude RICHARD ^b, Christian LOVIS ^b, Eirik ÅRSAND ^d, Jessica ROCHAT ^b, and Gunnar HARTVIGSEN ^a

^a UiT The Arctic University of Norway, Tromsø, Norway.

^b University of Geneva, Geneva, Switzerland.

^c Illinois Institute of Technology, Illinois, USA.

^d Norwegian Centre for E-health Research, Tromsø, Norway.

Abstract. The rapid improvement in mobile health technologies revolutionized what and how people can self-record and manage data. This massive amount of information accumulated by these technologies has potentially many applications beyond personal need, i.e. for public health. A challenge with collecting this data is to motivate people to share this data for the benefit of all. The purpose of this study is to survey and examine factors that may motivate sharing this data. We asked 447 participants four questions related to health data sharing and motivation. Participants with a chronic disease were concerned about data sharing but also willing to share health data if personalized feedback is provided. Functionality, ease of use, and privacy are regarded as crucial features of health apps.

Keywords. Public health, mHealth, motivation, health surveys, health data sharing.

1. Introduction

The ubiquitous nature of smartphones, wearables, and sensors have revolutionized the way people collect health-related data. An increasing number of people collects large amounts of data for disease-management, fitness, and self-surveillance. In a recent study, more than 60% of American participants tracked various health parameters, including diet, weight, and physical exercise, where 21% relied on fitness tracker technology [1]. People with chronic diseases use different technologies, e.g. mHealth apps and continuous glucose monitors for diabetes management, collecting and processing health data for their self-management. This data can potentially be used as a secondary source of information for public health, including tracking of disease trends, behavioural patterns over time, chronic diseases status, research, and policy work [2, 3, 4, 5]. The impact of these self-collected data highly depends on people's willingness to share their data for the intended purpose. Considering the potential of these data to inform about individual and population health, understanding the users' expectations and willingness towards mass data sharing is an important area of research. Various factors could affect people's motivations to engage in mass data sharing, e.g. lack of trust, which is mainly

¹ Corresponding author, E-mail: andre.henriksen@uit.no ; * Authors contributed equally

subject to data security, privacy, and confidentiality issues [6]. When sharing such data, expectation and willingness often differ in different patient groups, as well as in healthy people [7, 8], e.g. people with type 1 diabetes have a high need to record health data several times per day. Here, we examine factors related to people's knowledge and expectations toward raising motivation for sharing of health-related data and comparing these factors across different groups.

2. Method

We created an online survey with questions related to motivational factors around mHealth apps and data sharing. Questions were derived from 16 in-person interviews [6]. The survey was conducted among English-speaking internet users in a Swiss cohort of healthy people and also in English and Norwegian online diabetes groups. We collected data between 11/2018 and 08/2019. The questions related to data sharing are: 1) How concerned are participants about sharing health data, 2) what do participants expect in return when sharing health data, 3) which data types are participants willing to share anonymously for research, and 4) how important are different criteria in order to agree to install an application that collects and shares data from their wearable device. Options were on a 4-point Likert scale from "not important at all" to "very important" (including "I don't know"), or multiple-choice. We stratified responses into three groups: 1) People with diabetes (PWD), 2) people with other chronic disease and 3) people without a chronic disease. Details about the questionnaire are available at DataverseNO [9]. We report descriptive statistics on age, gender, experience using wearable devices and mobile health apps and wearables for sharing health or activity data.

3. Results

Four hundred forty-seven (447) participants finished the survey, of which eight did not answer whether they have a disease, and nine selected "Do not want to answer". Further analyses are based on the remaining $N=430$. Sixty-one (61) participants had diabetes, 82 participants had another chronic disease, and 285 had no chronic disease. The majority of participants came from Switzerland (187), Norway (59), US/UK/Australia/Canada (77), France (26), and Germany (13). Remaining 46 came from 35 countries covering all continents. Table 1 gives their demographics, familiarity with mHealth apps, and sharing experience. There was no age ($p=.083$) or gender ($p=.133$) difference between the groups. However, 97% in the diabetes group use a wearable device for collecting activity or other health data, compared to only 51% in the "no disease" group, and 55% in the "other chronic disease" group, $\chi^2(423)=44.04$, $p=.001$. Many PWD has experience in sharing data, less so people without a chronic disease, $\chi^2=19.6$, $p < .001$.

Table 1. Demographics and familiarity with mHealth apps. Na= no answer.

	With diabetes	With other chronic diseases	Without chronic disease
Age: <30 y; 30-50y; >50y	15; 17; 27 (Na: 2)	9; 32; 40 (Na: 1)	34; 96; 146 (Na: 6)
Gender: female; male; other	35; 25; 1	59; 20; 3	177; 99; 3
Wearable device: yes; no	59; 2	44; 36	143; 137
Sharing experience: yes; no	24; 27 (Na: 10)	22; 45 (Na: 14)	47; 160; (Na: 11)

3.1. Technologies use for health tracking

87% of those with diabetes use a health-specific device, in addition to often using a Physical Activity (PA) tracker. 3% use no sensor or wearable device. Among those with other- or no chronic diseases, the use of sensors integrated in the smartphone and PA trackers are most common. 25-30% use mobile health apps. Generally, participants preferred to discuss health issues with health providers. PWD would discuss it with others PWD (17%), but rarely with their family (6%). This is in contrast to persons with other- or no chronic disease, where 19% and 20%, respectively, would discuss health issues with their family and friends, see Figure 2b. This difference with whom to share was statistically significant, $\chi^2=34.67$, $p<.001$.

3.2. Concerns about data sharing

Regarding what people are most concerned about sharing health data, persons with no chronic disease are in general least concerned, and all three groups rate storage as least and transparency as most concerning. Figure 1a shows how each group rate concerns about confidence and trust, data ownership, storage location/availability, and transparency of third party usage. Figure 1b shows how important certain features are for each group in order to agree to install an application that collects and share health data from their wearable device.

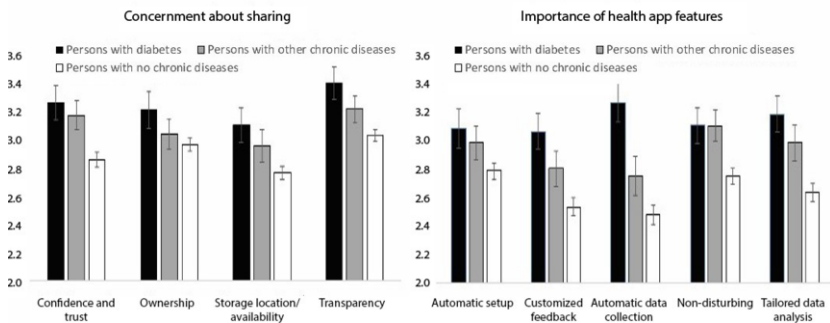


Figure 1. a: Concerns about different aspects of data sharing. b: Importance of features, answer option 1 to 4.

3.3. Participants expectations for sharing health data

Regarding what participants expected in return for sharing their health data, personalized feedback was chosen by 60% of participants, integrated view (i.e. aggregated results), was chosen by 53%, decision support by 36%, and least chosen (16%) was comparing

status with others. There was no difference between the groups in their expectations, smallest $p > .74$, but comparing with others was rated as least important (Figure 2a).

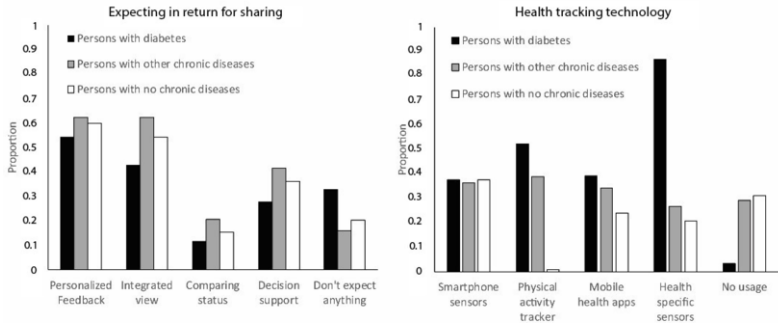


Figure 2. a: Expectations for sharing data. b: Group distribution of health tracking technology.

3.4. Willingness to anonymously share given data types in a research project

Groups do not differ in sharing lifestyle/dietary information, signs of infection, daily mood, geographical location, sleep duration, social environment (corrected for multiple comparisons). Participants are least willing to share their geographical location. Persons with chronic disease are more willing to share medication intake, physiological indicators and their weight. An overview is given in Table 2.

Table 2. Overview of willingness to share specific health data by group. Chr. = chronic.

	Medication intake and treatment	Lifestyle/ dietary intakes	Signs of infection	Physiological indicators	Daily mood	Physical activity	Geographical location	Sleep duration	Weight	Social environment	None of these
N	213	219	195	202	193	248	87	240	221	173	89
Diabetes	53	41	31	46	29	43	20	43	43	30	5
Other chr.	44	45	45	40	46	54	16	50	49	35	16
No chr.	116	133	119	116	118	151	51	147	129	108	68
X ²	43.48	9.05	5.22	24.39	5.73	8.93	6.92	8.29	15.48	2.87	7.59
p-value	<.001	.011	.074	<.001	.057	.012	.031	.016	<.001	.238	.023

3.5. Features' importance in mHealth app

Regarding which features of a health-app is important to participants (1=not important to 4=very important), non-disturbed tracking and automatic setup is rated by most as important, i.e. the main effect of feature, $F(3.83, 1163.75)=3.389$, $p=.01$, $\eta^2=.011$. All features were more important to people with a chronic disease, $F(2, 304)=12.09$, $p<.001$, $\eta^2=.074$. PWD rate automatic data collection as most important, yielding a significant interaction effect, $F(7.66, 1163.75)=2.104$, $p=.035$, $\eta^2=.014$.

4. Discussion

Mass sharing of health data could provide vital information for individual health management and public health. Continuous collection of quality health-related data and willingness to share these are limited by the user's motivation and expectations. The rate of acceptance of health tracking devices among clinicians is increasing [10], but

retention is decreasing [11]. Compared to PWD who use wearable sensors, participants with no- and other chronic disease reported higher adoption of sensors integrated in the smartphone and PA trackers. Automatic data collection, easier interface, e.g. voice command, tailored and personalized feedback is likely to increase usage and long-term engagement in such devices.

Concern about health data sharing is dependent on the type of data, and related concerns on issues such as privacy, security, confidentiality, transparency, and ownership. Participants with no chronic disease are less concerned compared to people with a chronic disease. Indeed, potential consequences of data leakage like repeated low blood glucose level (hypoglycemia) might result in the suspension of one's driving license or disqualification of health insurance enrollment in some countries. Some privacy and security shall be kept, i.e. many participants do not want to share their geographical location.

This study on how to motivate health data sharing is a collaboration between several projects where different systems for health data collection are under development. Results from the present study, and upcoming publications on related topics, will be used to direct the implementations of these systems for maximum acceptance. Future works include data collection, data quality and accuracy analysis, and detecting health patterns at the population level and in people with specific chronic diseases. Fulfilment of participant's expectations and resolving individual concerns could motivate sharing health-related data. Results indicate that participants expect some kind of immediate benefits from sharing their data, including tailored and personalized data analysis, integrated view, feedback and others. Comparison of status among peers was found to be less relevant.

References

- [1] J. Vitak, Y. Liao, P. Kumar, M. Zimmer, K. Kritikos (eds.), *Privacy Attitudes and Data Valuation Among Fitness Tracker Users*, Cham: Springer International Publishing, 2018.
- [2] J. Chen, A. Bauman, M. Allman-Farinelli, A Study to Determine the Most Popular Lifestyle Smartphone Applications and Willingness of the Public to Share Their Personal Data for Health Research, *Telemed J E Health* **22**(8) (2016), 655-65.
- [3] K.T. Pickard, M. Swan, Big Desire to Share Big Health Data: A Shift in Consumer Attitudes toward Personal Health Information, *2014 AAAI Spring Symposium Series*, 2014.
- [4] K. Ostherr, S. Borodina, R.C. Bracken, C. Lotterman, E. Storer, B. Williams, Trust and privacy in the context of user-generated health data, *Big Data & Society* **4**(1) (2017).
- [5] A. Seifert, M. Christen, M. Martin, Willingness of Older Adults to Share Mobile Health Data with Researchers, *GeroPsych* **31**(1) (2018), 41-9.
- [6] A.Z. Woldaregay, D.Z. Issom, A. Henriksen, H. Marttila, M. Mikalsen, et al., Motivational Factors for User Engagement with mHealth Apps, *Studies in health technology and informatics* **249** (2018), 151-7.
- [7] K.K. Kim, J.G. Joseph, L. Ohno-Machado, Comparison of consumers' views on electronic data sharing for healthcare and research, *J Am Med Inform Assoc.* **22**(4) (2015), 821-30.
- [8] D.J. Willison, V. Steeves, C. Charles, L. Lotterman, J. Ranford, et al., Consent for use of personal information for health research: do people with potentially stigmatizing health conditions and the general public differ in their opinions?, *BMC Med Ethics* **10** (2009), 10.
- [9] A. Henriksen, A.Z. Woldaregay, D-Z. Issom, G. Pfuhl, A. Richard, et al., *Questionnaire for motivation in mobile health*, DataverseNO V1, 2019. <https://doi.org/10.18710/28SRMJ>
- [10] M. Bradway, L. Ribu, G. Hartvigsen, Årsand E, The evolution of clinicians' preparedness for mHealth use (2013-2017) and current barriers, *Proceedings from the 16th Scandinavian Conference on Health Informatics 2018*, Linköping University Electronic Press, Denmark, 2018.
- [11] C. Attig, T. Franke, Abandonment of personal quantification: A review and empirical study investigating reasons for wearable activity tracking attrition, *Computers in Human Behavior* **102** (2020), 223-37.

Section 5

Human Factors and Citizen Centered Digital Health

This page intentionally left blank

A Path Towards Inclusion: Transdisciplinary Experience for the Inclusion of Self-Perceived Gender in an Information System

Melanie CASSARINO^{a,1}; Eunice CORREA^a; Sebastián MINOLETTI^a; Amalia BOTTO^a; Romina RAPISARDA^a; María Florencia GRANDE RATTI^a; Daniel LUNA^a

^a*Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina*

Abstract. In Argentina, since 2012 the Gender Identity Law guarantees that people who identify as transgender must be called by their chosen name, among other things. Transgender patients are usually not represented in information systems, whether electronic or not. This can place them in situations of vulnerability and discrimination. With that in mind, an interdisciplinary team was formed to take address the problems of this population regarding to the information systems of Hospital Italiano de Buenos Aires. The aim of this study is to spread the experience of the transdisciplinary team on the Electronic Health Record's modification and other associated applications to achieve include the self-perceived gender identity and the chosen name of those people who modified their gender.

Keywords. Transgender, Electronic Health Record, User-centered design, prototype

1. Introduction

In health systems it is common that people who identify with a different gender or name than the one that was assigned to their birth, suffer discrimination [1] by both administrative and healthcare staff. According to a survey conducted by INDEC and INADI in 2012 [2], three out of ten transgender or transvestite women present stories of abandonment of medical treatment due to discrimination in health institutions. For the same reason, five out of ten, stopped going to hospitals. The same study also indicates that a significant barrier to access the health system is the lack of coverage. It is estimated that 80% of this population does not have access to health insurance or prepaid medicine [3].

Gender identity is understood as “the internal and individual experience of gender as each person feels it, which may or may not correspond to the sex assigned at the time of birth, including the personal experience of the body. This may involve the change of the appearance or body function through pharmacological, surgery or other methods, when it is freely chosen. It also includes other gender expressions, such as dress, speech and manners” [4]

According to a study carried out by Fundación Huésped in 2011 [5] about the perceptions and experiences of stigma and discrimination in populations of trans

¹ Melanie Cassarino, [E-mail: melanie.cassarino@hospitalitaliano.org.ar](mailto:melanie.cassarino@hospitalitaliano.org.ar)

women, men who have sex with men and drug users, it was found that trans women were the most systematically excluded group from an early age and that their social welfare was below expectations due to stigma and its consequences, such as the legal non-recognition of their identity.

In 2012, Law 26743 of "Gender Identity" [4] is sanctioned in Argentina, which defines in its Article n°1 that every person has the right:

To the recognition of their gender identity;

To the free development of his person according to his gender identity;

To be treated accordingly to their gender identity and, in particular, to be identified in this way in the instruments that prove their identity regarding to the name, image and sex with which it is registered there.

The Ministry of Health of the Argentinian Republic, issued a "Guide for health teams" on comprehensive health care for transgender people [6]. This guide mentions that people must be respected in their name and gender of choice within the health system, both with respect to the physical place of hospitalization and in the registration instruments (medical history, certificates, records, etc.).

There's plenty bibliography about the inclusion of sex, gender and chosen name in in electronic health records [7][8][9][10][11][12][13][14]. Reviewing all of these papers, we could agree that there is a general consensus that these changes can positively affect this population's quality of care and medical attention.

1.1. Situation in Hospital Italiano de Buenos Aires

The Hospital Italiano de Buenos Aires (HIBA) it is a private hospital, of high complexity that has 780 beds, 40,000 hospitalizations per year and more than 2 million outpatient consultations. Within its structure is the Department of Health Informatics that develops the information systems within the institution. Since 1995, the Hospital has an electronic information system that covers both the assistance and administrative and management areas. The department is responsible both for the design of the tools and also for the storage's optimization and information retrieval with the aim of increasing patient safety, reducing error and increasing quality of care.

The HIBA has an application for the registration of patients. In 2017, with the objective to apply Law 26743 to the Master Patient Index (MPI), it was decided to develop a modification in it [15] with the aim of include the different identities that coexist in the Argentine territory as well as the use of the chosen name for people identification.

Analyzing the current bibliography and the sanctioned Argentine law, the possibility of registering the following items within the MPI was added:

- Chosen name
- Self-perceived gender identity (with structured fields: Man, Woman, Trans - man, Trans - woman, Intergender, Other)

After the implementation of the change in the registration system, the work team raised the need to include this information into the Electronic Health Record (EHR) so that the self-perceived gender identity and the chosen name are available to all healthcare professionals.

Because of that, the Department of Health Informatics is currently working on the visualization of this new information in the different available applications: from the patient caller system located in the waiting rooms, to Electronic Health Record, where

health professionals will have a comprehensive record that will allow them to provide holistic healthcare.

The aim of this study is to spread the experience of the transdisciplinary team on the EHR's modification and other associated applications to achieve the inclusion of self-perceived gender identity and the chosen name of those people who modified their gender.

2. Methods

In the meetings held with the healthcare team that take care of the trans population, the following needs were detected:

- View the name chosen as well as the name that appears on the National Identity Document (DNI).
- Known the person's biological sex and self-perceived gender.
- Identify which organs related to biological sex the person has [16].
- Identify what surgeries or hormonal treatments the person received [17].

The Health Informatics work team decided to divide these needs into two work stages;

First stage: within this first stage, we worked on the visualization of the people's identification information (chosen name, DNI name, biological sex and self-perceived gender) in the different applications.

Second stage: it was defined to focus the second stage of work on the needs related to clinical care: organs, surgeries and hormonal treatments of patients.

In this study we will focus mainly on the description of the work done during the first stage of the project.

Based on the need detected in the healthcare staff to be able to visualize the identification information of the people, it was decided to carry out the card sorting technique in a random way to different people from the HIBA (nurses, physicians, administrative staff). This technique is used to investigate how users think that information on digital products should be organized and / or categorized. Because the HIBA's EHR has iconography to represent the gender of the patients, for this work, the "card sorting" technique was used to identify the knowledge that the staff had about the icons that represent trans and intergender identity. After this process, we proceeded to choose the icon that represents the transgender population of Argentina to be included in the EHR.

Then, different prototypes of the EHR that contemplate the person's chosen name, the surname corresponding to the DNI and the initials of the DNI name in parentheses, were designed.

Structure: [First name + last name + (initials first name + initials last name)] (Figure 1)

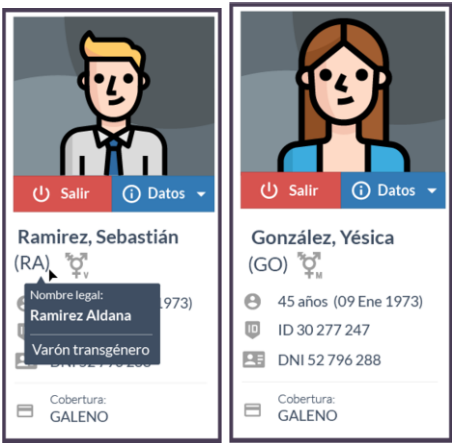


Figure 1. Representation in the EHR

These designs were prototyped and tested with end users who were unaware of this project. This process will be published in another research paper.

As part of the inquiries and while the design was being made for the EHR, a meeting was held with representatives of the Health Secretariat of the Argentine LGBT Federation (FALGBT) in order to showing them the prototype created for the EHR and the election of iconography for gender representation.

During this meeting the FALGBT provided advice on the visualization of information, privacy/confidentiality of data and information regard to trans and intergender identities. They considered that the presented prototype was adequate to represent transgender identities. Also, they suggested continuing with the second stage (detailed above) in order to integrate this information with the patient's clinical information.

3. Conclusions & Discussion

Throughout this process, we re-discovered the importance of teamwork and the added value that interdisciplinarity gives. Having user-centered design specialists and health professionals with informatics orientation within the team, allowed a holistic vision when devising the solution for the correct identification within the EHR.

Besides it was very important the opportunity to have meetings with health professionals at the time of making the inquiries, validating the information and being able to detect their needs to continue the path towards inclusion.

Our next step is to incorporate a list of surgeries, hormonal treatments and other clinical data relevant to the care of this population.

References

- [1] A. Tomicic, C. Gálvez, C. Quiroz, C. Martínez, J. Fontbona, J. Rodríguez, et al., Suicidio en poblaciones lesbiana, gay, bisexual y trans: revisión sistemática de una década de investigación (2004-2014). *Rev méd Chile*. 2016;144(6):723–33.
- [2] I. Indec, Primera Encuesta sobre Población Trans 2012: Travestis, Transexuales, Transgéneros y Hombres Trans. Informe Técnico de la Prueba Piloto. Municipio de la Matanza. 2012;
- [3] Akahatá, Agrupación Nacional Puntos Peronistas, Cooperativa de Trabajo La Paquito, Abogados y abogadas del NOA en Derechos Humanos y Estudios Sociales (ANDHES), Arte Trans, Asociación de Lucha por la Identidad Travesti y Transexual (ALITT), et al. [Internet]. 2016 [cited 2019 Sep 5]. Available from:
[1] https://tbinternet.ohchr.org/Treaties/CEDAW/Shared%20Documents/ARG/INT_CEDAW_NGO_ARG_25486_S.pdf
- [4] Ley n°26743 de Identidad de género [Internet]. [cited 2019 Jul 30]. Available from: https://www.buenosaires.gob.ar/sites/gcaba/files/ley_26.743_de_identidad_de_genero.pdf
- [5] I. Aristegui, V. Zalazar, Ley de identidad de género y acceso al cuidado de la salud de las personas trans en Argentina: resumen ejecutivo. Fundación Huésped; 2014.
- [6] M. de Salud Argentina, Atención de la salud integral de personas trans. Guía para equipos de salud., 2015.
- [7] J.D. Bosse, R.G. Leblanc, K. Jackman & R.I. Bjarnadottir, Benefits of Implementing and Improving Collection of Sexual Orientation and Gender Identity Data in Electronic Health Records. *Computers, Informatics, Nursing: CIN* (2018). 36(6), 267–274.
- [8] S. Cahill, and H. Makadon, Sexual Orientation and Gender Identity Data Collection in Clinical Settings and in Electronic Health Records: A Key to Ending LGBT Health Disparities. *LGBT Health*, (2014). 1(1), 34–41.
- [9] S.R. Cahill, K. Baker, M.B. Deutsch, J. Keatley, and H.J. Makadon, Inclusion of Sexual Orientation and Gender Identity in Stage 3 Meaningful Use Guidelines: A Huge Step Forward for LGBT Health. *LGBT Health*, (2016). 3(2), 100–102.
- [10] M.B. Deutsch, and D. Buchholz, Electronic health records and transgender patients--practical recommendations for the collection of gender identity data. *Journal of General Internal Medicine*, 2015. 30(6), 843–847.
- [11] M.B. Deutsch, J. Green, J. Keatley, G. Mayer, J. Hastings, A.M. Hall, and World Professional Association for Transgender Health EMR Working Group, Electronic medical records and the transgender patient: recommendations from the World Professional Association for Transgender Health EMR Working Group. *Journal of the American Medical Informatics Association: JAMIA*, (2013). 20(4), 700–703.
- [12] C. Grasso, M.J. McDowell, H. Goldhammer, and A.S. Keuroghlian, Planning and implementing sexual orientation and gender identity data collection in electronic health records. *Journal of the American Medical Informatics Association: JAMIA*, (2019). 26(1), 66–70.
- [13] Institute of Medicine, and Board on the Health of Select Populations, Collecting Sexual Orientation and Gender Identity Data in Electronic Health Records: Workshop Summary. National Academies Press, 2013.
- [14] A.M. Shenoy, N. Kashey, P. Brown, and N.R. Kudler, Better representing transgender patients in the electronic health record. *Journal of Health Care for the Poor and Underserved*, (2014). 25(4), 1641–1645.
- [15] J. Frangella, M. Cassarino, F. Plazzotta, F. Gassino, C. Otero, D. Luna, Designed Strategies and Adaptation of a Master Patient Index for Transgender Patients in a Tertiary Care Hospital. *Stud Health Technol Inform.*, 2019 Aug 21;264:1698–9.
- [16] Deutsch MB, Green J, Keatley J, Mayer G, Hastings J, Hall AM, et al. Electronic medical records and the transgender patient: recommendations from the World Professional Association for Transgender Health EMR Working Group. *J Am Med Inform Assoc*. (2013) 20(4):700–3.
- [17] R. Abdala, A. Nagelberg, F. Silveira, P. Otero, E. Mormandi, Perfil de seguridad a corto plazo de la terapia hormonal cruzada en trans-varones. *Medicina*. (2018). 78(6):399–402.

A Systematic Review on CDSS Alert Appropriateness

Olufisayo OLAKOTAN^{a,1}, Maryati MOHD YUSOF^b and Sharifa EZAT WAN PUTEH^c

^{a, b} *Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia*

^c *Faculty of Medicine, Univerisiti Kebaangsaan Malaysia*

Abstract. Clinical decision support systems (CDSSs) provides vital information for managing patients by advising clinicians through an alert or reminders about adverse events and medication errors. Clinicians receive a high number of alerts, resulting in alert override and workflow disruptions. A systematic review was carried out to identify factors affecting CDSS alert appropriateness in supporting clinical workflows using a recently introduced framework. The review findings identified several influencing factors of CDSS alert appropriateness including: technology (usability, alert presentation, workload and data entry), human (training, knowledge and skills, attitude and behavior), organization (rules and regulation, privacy and security) and process (waste, delay, tuning and optimization). The findings can be used to guide the design of CDSS alert and minimise potential safety hazards associated with CDSS use.

Keywords. CDSS, Alert fatigue, Alert override, Framework, Appropriateness, Clinical workflow

1. Introduction

Clinical decision support systems (CDSS) has widely been implemented to improve care quality, patient safety, reduce costs and enhance healthcare outcomes [1]. CDSS encompasses various tools such as automated alerts and reminders that provide its users with information that can enhance clinical decision making [2] through evidence-based recommendations to clinicians at the point of care [3]. Despite the wide adoption of CDSS, clinicians receive numerous alerts, leading to a phenomenon known as ‘alert fatigue’ [4]. These alerts are meant to avoid possible safety risks, such as drug-drug interactions (DDI) and allergies.

However, clinicians override relevant and irrelevant alerts, which may lead to adverse events or unintended patient harm. A significant reason for alert override is disruption in clinician’s workflow [5]. Therefore, evaluating the appropriateness of CDSS alerts and its effect on workflow enables us to understand how alerts can be triggered effectively to support clinical workflow. This paper presents a systematic review on factors affecting CDSS alert appropriateness in supporting clinical workflow.

¹ Corresponding Author, E-mail: P84271@siswa.ukm.edu.my.

2. Theoretical background

The new framework for evaluating CDSS alert appropriateness in clinical workflow is proposed based on critical review and adaptation from previous health information technology framework including Human-Organization-Technology-fit (HOT-fit) [6,7], Human Automation and Interaction (HAI), [8], business process management (BPM) life cycle [9], and Systems Engineering Initiative for Patient Safety (SEIPS) [10] (Figure 1).

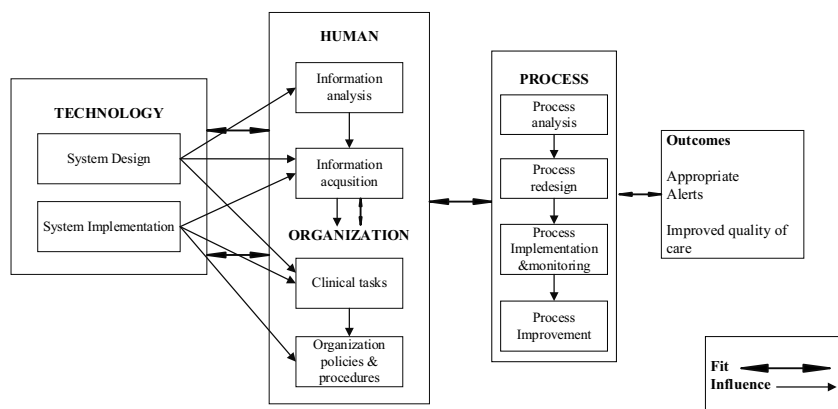


Figure 1. CDSS appropriateness framework

3. Methodology

This review seeks to answer “what are the factors affecting CDSS alert appropriateness in supporting clinical workflow?”. The review followed the guidelines of Preferred Reporting Items for Systematic Reviews (PRISMA). Bibliographic databases namely PubMed, Scopus, ACM, Science Direct, IEEE, Ovid Medline, and Ebscohost, were searched for relevant literature from 1997 to 2018. The search was conducted on October 11, 2018 to February 28, 2019. The following keywords such as clinical decision support system (CDSS), alert fatigue, alert override, electronic health record (EHR), electronic medical record (EMR), health information system (HIS), were used to search the databases. We included studies on features and functions pertinent to CDSS alert appropriateness and alert and workflow related problems in healthcare settings. Studies on medical devices with alarms and beeps, such as physiological monitors, ventilators, and pumps, were excluded. A quality assessment tool known as Critical Skill Appraisal Program (CASP) for qualitative, quantitative and retrospective study was used to evaluate the quality of the selected articles.

The bibliographic databases returned 9769 publications. Then, 8899 titles and abstracts were initially screened, and 7796 were excluded. In summary, 76 papers were obtained (64 main papers and 12 additional papers from citation chaining). Deductive data analysis was conducted based on the proposed evaluation framework dimensions and measures, whereas inductive analysis enabled us to identify new themes and sub-themes such as knowledge base, alert flexibility, alert content and workload which are categorized under system design and implementation dimensions.

4. Results

Table 1 highlights key factors affecting CDSS alert appropriateness in supporting clinical workflow.

Table 1. Factors affecting CDSS alert appropriateness in supporting clinical workflow

Technology	Human	Organization	Process
System design	Information acquisition	Policy and procedure	Process analysis
Alert content (N = 14)	Knowledge and skills (N = 5)	Privacy and security (N = 5)	Waste (delay (N = 6); others (N = 6))
specificity/ sensitivity (N = 7)	Training (N = 5)	Rules and regulations (N = 3)	Process redesign
Usability (N = 7)	Information analysis		Modification (N = 7)
Alert presentation (N = 7)	Perception (N = 7)		Process mapping (N = 1)
Alert flexibility (N = 7)	Behavior (N = 6)		Process implementation
Knowledge base (N = 5)	Attitude (N = 3)		Automation (N = 9)
System implementation			Process monitoring/ improvement
Workload (N = 8)			Bottleneck (N = 7)
Hardware and software (N = 7)			Tuning and optimization (N = 4)
Documentation (N = 6)			
Communication (N = 6)			

Technology. Usability problems such as illegible screens, excessive mouse clicks and poorly designed drop-down menus that are counter-intuitively arranged, disrupts clinician’s workflow [11,12]. Moreover, alert fatigue and override persists because alerts do not indicate severity levels (i.e., mild, moderate and severe) [12]. Alerts should contain reliable and clear text messages that can inform or enhance clinician’s decision making. Studies reports that knowledgebase problems such as poorly categorized codes and wrongly configured default settings contributes to adverse events and medication errors [13,14]. Therefore, CDSS rules, codes, and default settings must be properly modified to present accurate and up-to-date information to users during patient care [14]. Poorly implemented CDSS has created additional problems for busy clinicians such as workload, data entry burden, system downtimes, hardware and software breakdown [1,14]. For instance, system is down when the hardware and software cannot handle transaction loads [1]. Software malfunctions such as outdated versions of operating systems and system lockout hindered clinicians from entering and verifying orders [14,15]. Successful implementation of CDSS requires a proper understanding of clinician workflow to ensure the fit between them.

Human. Inadequate knowledge and skills hindered clinicians from using CDSS with its alert function effectively [1,16]. For example, incompetent clinicians cannot create, view, and adjust reminders (e.g., removing inapplicable reminders). Clinicians without any formal training or education on how to use CDSS might engage in self-learning, which leads to alerts misinterpretation [1,15]. However, most clinicians agree that hands-on training will improve their skills in handling CDSS [14,16]. Clinician’s attitude towards alerts depends on how alerts are displayed (i.e. intrusive or interruptive) during workflow [17]. CDSS alert that provides irrelevant information can subconsciously change prescriber’s behavior during prescription [5].

Organization. Flawed organizational policies and procedures can contribute to CDSS failure and workflow disruption. Hospital rules and regulations resulted in rigid task execution such as inability to adjust alert settings [17], tedious process of overriding password and justifying alert override prior ordering [18]; inability to access most CDSS features due to the security measures [19]; conflicting alerts; [20]; and silenced low-level alerts for drug-drug interactions (DDI) [21].

Process. Delays caused by system downtime impedes patient flow and disrupts clinician's workflow [22]. Examples include a lockout system feature that hindered clinician from ordering medication, resulting delays in prescribing medication; switching between screens and overriding alert caused non-value added activities (i.e. waste); [23]. Monitoring recurring errors and identifying bottlenecks such as alert fatigue, alert overload during the clinical process will minimize workflow disruption [13]. In addition, tuning and optimization strategies such as the use of alert evaluation dashboard to monitor alert override rates or physician response to alerts will prevent recurring errors and subsequently enhance clinician's workflow [23].

5. Discussion and conclusions

This review evaluated CDSS alert appropriateness that was mostly contributed by technology factors, particularly system design that is related to alert content, accuracy, and usability. Process factors are the second highly influencing factors where waste, including delay, and bottleneck from the use of CDSS alert required process modification and redesign to improve workflow efficiency. Human factors reported negative user perception and behavior towards CDSS alerts that are highly influenced by poor system design, followed by inadequate skills and training. Disruptive CDSS alerts that affect clinical processes were highly reported. Therefore, CDSS must be well designed to support clinical tasks, workflow, and cognitive activities and aligned with the socio-technical aspects. Organizational policies and procedures should be flexibly adapted to accommodate CDSS implementation. Continuous process improvement methods such as BPM and Lean can be implemented to identify and eliminate inefficient clinical processes impeding the fit between technology, human, organization. Our findings indicate the usefulness of the proposed evaluation framework in addressing alerts and workflow problems in clinical settings. Future studies should address cognitive work challenges or activities that impede the use of CDSS alerts.

Acknowledgement

We gratefully acknowledge the funding from Ministry of Higher Learning Malaysia (FRGS/1/2018/ICT04/UKM/02/5) for sponsoring this study.

References

- [1] M. Lugtenberg, J.W. Weenink, T. Van Der Weijden, G.P. Westert, R.B. Kool, Implementation of multiple-domain covering computerized decision support systems in primary care: A focus group study on perceived barriers, *BMC Med. Inform. Decis. Mak.* **15** (2015) 82.
- [2] N.S. Bauer, A.E. Carroll, C. Saha, S.M. Downs, Experience with decision support system and comfort with topic predict clinicians' responses to alerts and reminders, *J. Am. Med. Informatics Assoc.* **23** (2016) e125–e130.

- [3] D.S. McEvoy, D.F. Sittig, T.-T. Hickman, S. Aaron, A. Ai, M. Amato, D.W. Bauer, G.M. Fraser, J. Harper, A. Kennemer, M.A. Krall, C.U. Lehmann, S. Malhotra, D.R. Murphy, B. O'Kelley, L. Samal, R. Schreiber, H. Singh, E.J. Thomas, C. V. Vartian, J. Westmorland, A.B. McCoy, A. Wright, Variation in high-priority drug-drug interaction alerts across institutions and electronic health records., *J. Am. Med. Inform. Assoc.* **24** (2017) 331-338.
- [4] J.J. Coleman, H. van der Sijs, W.E. Haefeli, S.P. Slight, S.E. McDowell, H.M. Seidling, B. Eiermann, J. Aarts, E. Ammenwerth, A. Slee, R.E. Ferner, R.E. Ferner, A. Slee, On the alert: future priorities for alerts in clinical decision support for computerized physician order entry identified from a European workshop., *BMC Med. Inform. Decis. Mak.* **13** (2013) 111.
- [5] M.T. Baysari, A. Tariq, R.O. Day, J.I. Westbrook, Alert override as a habitual behavior – a new perspective on a persistent problem, *J. Am. Med. Informatics Assoc.* **24** (2016) 409-412.
- [6] M.M. Yusof, J. Kuljis, A. Papazafeiropoulou, L.K. Stergioulas, An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit), *Int. J. Med. Inform.* **77** (2008) 386-398.
- [7] M.M. Yusof, A case study evaluation of a Critical Care Information System adoption using the socio-technical and fit approach, *Int. J. Med. Inform.* **84** (2015) 486-499.
- [8] R. Parasuraman, T.B. Sheridan, C.D. Wickens, A model for types and levels of human interaction with automation, *IEEE Trans. Syst. Man, Cybern.* **30** (2000) 286-297.
- [9] M. Dumas, M. La Rosa, J. Mendling, H.A. Reijers, Fundamentals of Business Process Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [10] P. Carayon, A. Schoofs Hundt, B.-T. Karsh, A.P. Gurses, C.J. Alvarado, M. Smith, P. Flatley Brennan, Work system design for patient safety: the SEIPS model., *Qual. Saf. Health Care.* **1** (2006) i50-8.
- [11] A. Feldstein, S.R. Simon, J. Schneider, M. Krall, D. Laferriere, D.H. Smith, D.F. Sittig, S.B. Soumerai, How to Design Computerized Alerts to Ensure Safe Prescribing Practices, *Jt. Comm. J. Qual. Saf.* **30** (2004) 602-613.
- [12] M.A. Krall, D.F. Sittig, Clinician's assessments of outpatient electronic medical record alert and reminder usability and usefulness requirements., *Proceedings. AMIA Symp.* (2002) 400-4.
- [13] S.Z. Kassakian, T.R. Yackel, P.N. Gorman, D.A. Dorr, Clinical decisions support malfunctions in a commercial electronic health record, *Appl Clin Inf.* **8** (2017) 910-923.
- [14] A. Wright, T.-T.T. Hickman, D. McEvoy, S. Aaron, A. Ai, J.M. Andersen, S. Hussain, R. Ramoni, J. Fiskio, D.F. Sittig, D.W. Bates, Analysis of clinical decision support system malfunctions: a case series and survey, *J. Am. Med. Informatics Assoc.* **23** (2016) 1068-1076.
- [15] R.J. Holden, What Stands in the Way of Technology-Mediated Patient Safety Improvements? A Study of Facilitators and Barriers to Physicians' Use of Electronic Health Records, *J. Patient Saf.* **7** (2011) 193-203.
- [16] J.J. Saleem, E.S. Patterson, L. Militello, M.L. Render, G. Orshansky, S.M. Asch, Exploring Barriers and Facilitators to the Use of Computerized Clinical Reminders, *J. Am. Med. Informatics Assoc.* **12** (2005) 438-447.
- [17] E.S. Patterson, B.N. Doebbeling, C.H. Fung, L. Militello, S. Anders, S.M. Asch, Identifying barriers to the effective use of clinical reminders: bootstrapping multiple methods., *J. Biomed. Inform.* **38** (2005) 189-99.
- [18] M.-L. Yeh, Y.-J. Chang, P.-Y. Wang, Y.-C. (Jack) Li, C.-Y. Hsu, Physicians' responses to computerized drug-drug interaction alerts for outpatients, *Comput. Methods Programs Biomed.* **111** (2013) 17-25.
- [19] A.E. Carroll, V. Anand, S.M. Downs, Understanding Why Clinicians Answer or Ignore Clinical Decision Support Prompts, *Appl. Clin. Inform.* **03** (2012) 309-317.
- [20] P. Zazove, M. McKee, L. Schleicher, L. Green, P. Kileny, M. Rapai, E. Mulhem, To act or not to act: responses to electronic health record prompts by family medicine clinicians., *J. Am. Med. Inform. Assoc.* **24** (2017) 275-280.
- [21] Y. Zenziper Straichman, D. Kurnik, I. Matok, H. Halkin, N. Markovits, A. Ziv, A. Shamiss, R. Loebstein, Prescriber response to computerized drug alerts for electronic prescriptions among hospitalized patients, *Int. J. Med. Inform.* **107** (2017) 70-75.
- [22] A.L. Russ, A.J. Zillich, M.S. McManus, B.N. Doebbeling, J.J. Saleem, A human factors investigation of medication alerts: barriers to prescriber decision-making and clinical workflow, *AMIA Symp.* (2009) 548-52.
- [23] M.H. Trivedi, E.J. Daly, J.K. Kern, B.D. Grannemann, P. Sunderajan, C.A. Claassen, Barriers to implementation of a computerized decision support system for depression: an observational report on lessons learned in clinical settings., *BMC Med. Inform. Decis. Mak.* **9** (2009) 6.

Analysing Sentiment and Topics Related to Multiple Sclerosis on Twitter

Guido GIUNTI^{a,1}, Maëlick CLAES^a, Enrique DORRONZORO ZUBIETE^b, Octavio RIVERA-ROMERO^b, Elia GABARRON^c

^aUniversity of Oulu, Oulu, Finland

^bUniversidad de Sevilla, Seville, Spain

^cNorwegian Centre for E-health Research, University Hospital North Norway, Tromsø, Norway

Abstract. *Background and objective:* Social media could be valuable tools to support people with multiple sclerosis (MS). There is little evidence on the MS-related topics that are discussed on social media, and the sentiment linked to these topics. The objective of this work is to identify the MS-related main topics discussed on Twitter, and the sentiment linked to them. *Methods:* Tweets dealing with MS in the English language were extracted. Latent-Dirichlet Allocation (LDA) was used to identify the main topics discussed in these tweets. Iterative inductive process was used to group the tweets into recurrent topics. The sentiment analysis of these tweets was performed using SentiStrength. *Results:* LDA identified topics were grouped into 4 categories, tweets dealing with: related chronic conditions; condition burden; disease-modifying drugs; and awareness-raising. Tweets on condition burden and related chronic conditions were the most negative ($p < 0.001$). A significant lower positive sentiment was found for both tweets dealing with disease-modifying drugs, condition burden, and related chronic conditions ($p < 0.001$). Only tweets on awareness-raising were most positive than the average ($p < 0.001$). *Discussion:* The use of both tools to identify the main discussed topics on social media and to analyse the sentiment of these topics, increases the knowledge of the themes that could represent the bigger burden for persons affected with MS. This knowledge can help to improve support and therapeutic approaches addressed to them.

Keywords. Multiple sclerosis, Twitter, natural language processing, sentiment analysis, topic modelling

1. Introduction

Multiple Sclerosis (MS) is the leading cause of non-traumatic neurological disability in young adults [1]. In order to successfully self-manage such a chronic condition, patients require to learn about and manage their symptoms and problems [2-4]. Moreover, they also need social support that enhances the awareness that they are cared for.

Social support is known to be an important aspect for persons with MS (pwMS) [5-7]. However, problems in societal participation, such as reduced interactions with family or friends, are common among people affected by this condition due to physical

¹ Corresponding Author: Guido Giunti E-mail: Guido.Giunti@oulu.fi.

disabilities and MS fatigue [8]. In this sense, the ubiquity of social media and their great potential to engage and communicate with others could make these channels valuable tools to support pwMS [9-10]. PwMS using social media could benefit from communities that share content linked to positive sentiment, as positive emotion motivates cooperation [11]. However, there is not much evidence on what are the topics being discussed on social media related to MS, and if these topics are linked to positive sentiment, and therefore could provide the needed social support.

The objective of this work is to identify the main topics related to multiple sclerosis that are discussed on Twitter, and to analyse the sentiment polarity linked to these topics.

2. Methods

We collected tweets from Twitter using the Twitter API matching the query: #ms OR #multiplesclerosis OR "multiple sclerosis" that were posted between February 9th 2019 and June 26th 2019, and were in English language.

In order to identify the topics discussed in these tweets, we utilized Latent-Dirichlet Allocation (LDA) [12]. LDA is a generative statistical model that assigns to each document (e.g., a single tweet) a probability to belong to each topic of a list of topics which number is fixed and usually small [13]. We pre-processed each tweet with the following steps: 1) All URLs were removed using the following regular expression: `https?://[[:alnum:]][:punct:]]*`; 2) All tweets were tokenized²; 3) All tokens were lemmatized using the R packages *textstem*³ and *lexicon*⁴; and 4) Tokens were filtered to remove stop words using stopword lists from R packages *lexicon* and *tm* [14] (English stopwords).

We built a document-term matrix from the lemmatized and filtered tokens of all tweets collected. Then we trained an LDA model using the 47,852 original English tweets posted between February 9th and May 2nd 2019 and 20 topics. We fine-tuned LDA hyper-parameters using a differential evolution algorithm [15] with R package *DEoptim* [16] with 200 iterations and a population size of 30. For each population member (a pair of values for hyperparameters) considered, an LDA model was generated using a random sample of 80% of the training data. The model was then evaluated computing the perplexity of the other 20% of the training data. At the end, we obtained hyperparameter values 0.292217 and 0.026492 that were used for the final model.

The sentiment analysis of the tweets was performed on the whole dataset (74,076 tweets) using SentiStrength [17]. This software is specially designed to analyse the sentiment in short texts by assigning two scores: the intensity of the positive sentiment (1 to 5) and the intensity of the negative sentiment (-1 to -5). To apply SentiStrength, all tweets were allocated in plain text file, which means that unicode special characters were removed from the original text. As a result of the analysis every tweet was assigned a positive and negative sentiment score.

² <https://cran.r-project.org/web/packages/text2vec/index.html>

³ <https://cran.r-project.org/web/packages/textstem/index.html>

⁴ <https://cran.r-project.org/web/packages/lexicon/index.html>

The treatment of personal information for this study was approved by the data-protection officer at the University Hospital of North Norway (Ref:02275).

3. Results

The API extracted a total of 74,076 original tweets (i.e., not retweeted) on MS written in English. The sentiment analysis showed an average positive strength of 1.6 (SD=0.8); and a negative strength of -1.78 (SD=1.0) on the extracted MS tweets.

The 20 topics identified with LDA were analysed and recurring themes were identified and grouped into 4 main topics through an iterative inductive process. These 4 topics dealt with: Related chronic conditions; Condition burden; Disease modifying drugs; and Awareness raising. Table 1 summarizes the main keywords identified with LDA and included in each of the topics.

Table 1. Recurring MS’ topics identified with LDA and grouped through iterative inductive process

Topic	Keywords included
Related chronic conditions	Autoimmune; autoimmune_disease; parkinsons; alzheimers; rheumatoid; rheumatoid_arthritis; chronic_illness; crwriter; neurodegenerative; mental; Gofundme; injury; nervous; nervous_system; spinal; central; spinal_cord; central_nervous; cord; brain_injury
Condition burden	Sign; ms_symptom; symptom_ms; quality_life; quality; warn_sign; sign_ms; sign_symptom; early_warn; risk_multiple; Cbd; cancer; marijuana; pain; dying; chronic_pain; bill; spasm; plea; cancer_multiple; Diagnose_multiple; diagnose; letsbeatms; intezaarnakaro_letsbeatms; inzaarnakaro; visit_intezaarnakaro; suffer_multiple; income; disability_progression
Disease modifying drugs	Cell; immune; sclerosis_patient; therapy; stem; trial; stem_cell; immune_system; disease_modify; phase; Drug; relapse; fda; approve; fda_approve; remit; treat_multiple; relapse_remit; oral; novartis
Awareness raising	Goal; goal_multiple; blog; count_donation; sclerosis_count; donation_donation; raise_goal; donation; count; energy; Month; awareness_month; march; ms_awareness; sclerosis_awareness; awareness; awareness_week; hob; ov; march_multiple

When comparing the sentiment polarity of tweets, we find that those with a lower sentiment polarity are the ones with a probability >0.6 to belong to the topics “disease modifying drugs” and “condition burden” (1.40 and 1.44 respectively, vs 1.60 for all other tweets, t-test p<0.01). Only tweets on awareness raising had a higher positive strength than the average (1.80 vs 1.60, t-test p<0.01). On the other hand, the tweets with a more negative strength were the ones dealing with condition burden and related chronic conditions (-2.53 and -2.14 respectively, vs -1.78, t-test p<0.001). Only tweets on awareness raising and disease modifying drugs had lower negative strength than the average (-1.23 and -1.61, vs -1.78, t-test p<0.05). Table 2 shows the average sentiment positive and negative sentiment values of the tweets with a probability >0.60 to belong to one of the 4 identified LDA topics, (t-tests compare the sentiment of these tweets with the tweets within the same topic with a probability <0.60).

Table 2. Average sentiment of tweets with a probability >0.60 to belong to each topic

LDA topic	Number of tweets	Positive Mean (SD)	Negative Mean (SD)
Related chronic conditions	1106	1.58 (0.6)*	-2.14 (1.3)*
Condition burden	2256	1.44 (0.6)*	-2.53 (1.3)*
Disease modifying drugs	1749	1.40 (0.6)*	-1.61 (0.7)*
Awareness raising	1171	1.80 (0.7)*	-1.23 (0.6)*
ALL TWEETS	74076	1.60 (0.8)	-1.78 (1.0)

* Students t-test, p<0.01

4. Discussion

The MS topics discussed on Twitter that were identified with LDA, and that we grouped into four main categories, are consistent with the topics discussed in a recent publication on MS Facebook groups [18]. The importance of information and awareness, treatment related issues, and other MS-related symptoms was also key in that study [18].

It is well-known that people affected with chronic conditions, such as MS, carry painful emotions in need of sharing [19], and that expressing these emotions is linked to feelings of relief and alleviation [20]. In our study we found that Tweets dealing with disease-modifying drugs, condition burden, and related chronic conditions had a significant lower positive sentiment than the average of Tweets on MS; tweets on condition burden and related chronic conditions were the most negative ones. These could represent the topics with more emotional burden for those affected with MS, and therefore the topics for which they should receive greater social support. This is in line with other literature on reported feelings of pwMS[5].

Strengths and limitations: Our study has several limitations. We extracted tweets using the standard Twitter API, that did not index all tweets published in the considered dates, and therefore, the data extraction might not be representative, and should be considered as a random sample of tweets on MS. Relevant content may not have been included due to our search strategy. Additional research could consider including other keywords related to MS, expand the search by including other languages, include other social media channels, analyse emojis and metadata such as users’ profiles. We used SentiStrength software to conduct the sentiment analysis of the collected tweets on MS. This software has not yet been validated for use in the health domain. However, it was developed using posts on Myspace and shows a good performance with informal text, such the ones that can be found on Twitter.

Conclusions: The use of tools to identify main discussed topics on social media, and to analyse the sentiment of these topics increases the knowledge of the themes that could be representing the bigger burden for people with MS. This knowledge can help to improve support and therapeutic approaches addressed to people with multiple sclerosis.

Acknowledgments

EDZ receives funding and is supported by the V Plan Propio de Investigación de la Universidad de Sevilla, Spain.

References

- [1] A. Compston, A. Coles, *Multiple sclerosis*. Lancet 2002;359(9313):1221-1231.
- [2] Schulman-Green D, Jaser S, Martin F, et al. *Processes of Self-Management in Chronic Illness*. J. Nurs. Scholarsh. 2012;44(2):136-144.
- [3] M. Marziniak, G. Brichetto, P. Feys, et al. *The Use of Digital and Remote Communication Technologies as a Tool for Multiple Sclerosis Management: Narrative Review*. JMIR Rehabil Assist Technol 2018;5(1):e5.
- [4] D. Kantor, J.R Bright, J. Burtchell. *Perspectives from the Patient and the Healthcare Professional in Multiple Sclerosis: Social Media and Patient Education*. Neurol Ther 2018;7(1), 23–36.
- [5] G. Giunti, J. Kool, O. Rivera-Romero, et al. *Exploring the specific needs of persons with multiple sclerosis for mHealth solutions for physical activity: a mixed-methods study*. JMIR mHealth uHealth 2018;6(2):e37.
- [6] D. Kantor, J.R Bright, J. Burtchell. *Perspectives from the Patient and the Healthcare Professional in Multiple Sclerosis: Social Media and Participatory Medicine*. Neurol Ther 2018;7(1), 37–49.
- [7] L. Lavorgna, D. Ippolito, S. Esposito, et al. *A disease in the age of the web: How to help people with Multiple Sclerosis in social media interaction*. Mult Scler Relat Disord 2017;17:238-239.
- [8] A.L Katz, T.J Braley, E. Foxen-Craft, et al. *How Do Pain, Fatigue, Depressive, and Cognitive Symptoms Relate to Well-Being and Social and Physical Functioning in the Daily Lives*. Arch Phys Med Rehabil 2017;98(11):2160-2166.
- [9] A. Arguel, O. Perez-Concha, L. SYW, et al. *Theoretical approaches of online social network interventions and implications for behavioral change: a systematic review*. J Eval Clin Pract 2018;24(1):212-221.
- [10] M.S.C Lim, C.J.C Wright, E.R Carrotte, et al. *Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites*. Health Promot J Austr 2016;27(3):187-197.
- [11] D.G Rand, G. Kraft-Todd, J. Gruber. *The Collective Benefits of Feeling Good and Letting Go: Positive Emotion and (dis)Inhibition Interact to Predict Cooperative Behavior*. PLoS One 2015;10(1):e0117426.
- [12] D.M Blei, A.Y Ng, M.I Jordan. *Latent dirichlet allocation*. Journal of machine Learning research 2003;3(Jan):993-1022.
- [13] C. Montenegro, C. Ligutom, III, J. Vincent Orio, and D.A.M Ramacho. 2018. Using Latent Dirichlet Allocation for Topic Modeling and Document Clustering of Dumaguete City Twitter Dataset. In Proceedings of the 2018 International Conference on Computing and Data Engineering (ICCDE 2018). ACM, New York, NY, USA, 1-5. DOI: <https://doi.org/10.1145/3219788.3219799>
- [14] I. Feinerer, K. Hornik, D. Meyer. *Text Mining Infrastructure in R*. Journal of Statistical Software 2008;25(5):1-54.
- [15] K.V Price, R.M Storn, J.A Lampinen (2006). *Differential Evolution - A Practical Approach to Global Optimization*. Berlin Heidelberg: Springer-Verlag.
- [16] K. Mullen, D. Ardia, D. Gil, et al., 'DEoptim': An R Package for Global Optimization by Differential Evolution. Journal of Statistical Software 2011;40(6):1-26.
- [17] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas. *Sentiment strength detection in short informal text*. J Am Soc Inf Sci 2010 Dec 15;61(12):2544-2558. [doi: 10.1002/asi.21416]
- [18] S. D Rosa, F Sen. *Health Topics on Facebook Groups: Content Analysis of Posts in Multiple Sclerosis Communities*. Interact J Med Res 2019;8(1):e10146.
- [19] B. Rimé. *Comment: Social Integration and Health: Contributions of the Social Sharing of Emotion at the Individual, the Interpersonal, and the Collective Level*. Emotion Review 2018;10(1):67-70.
- [20] J.A Holyst (Ed.) *Cyberemotions: Collective emotions in cyberspace*. Springer 2016.

Automatic Interpretation of Laboratory Tests and Its Influence on Follow-up

Georgy KOPANITSA^{a, 1}

^aITMO University, Saint Petersburg, Russia

Abstract. Failing to follow up on the abnormal test results can cause serious health problems to patients. We conducted a retrospective medical record review of 3200 randomly selected patients aged 18 to 76 in 14 state clinics and two private laboratory services querying the common regional patient registry. One patient could be included (1 clinical case) in the study only once. We invited patients to take part in the interviews to gain a deeper understanding of the motives to follow up or not after receiving a recommendation and explanation of the role of the automatically generated interpretation in this decision. A qualitative study of the patients' motivation was performed with a group of 689 patients. All the patients who received their interpretations showed a much higher follow-up rate (68% average) than the patients who did not receive interpretations (49 % average). The results of our research demonstrated that there is a significant impact on the patients' decision to follow up on the tests. Patients consider time factor as an important advantage of the computer interpretations and are willing to get automatic interpretations if they can receive it faster than the ones from their doctor (question 4: median = 3 out of 7). Discussing the reasons behind the decision to follow up, the patients do trust the computerized clinical decision support systems (question 5: median = 5 out of 7), however, they prefer to receive interpretations and recommendations from doctors (question 3: median = 7).

Keywords. laboratory test, interpretations, decision support systems, follow-up

1. Introduction

Today, the use of computerized clinical decision support systems (CDSS) is gaining its traction in helping patients in various clinical situations [9; 19]. Such systems can produce patient-specific reminders, interpretations, and recommendations motivating a more appropriate delivery of care [10; 14]. For example, such systems are used to interpret laboratory test results [16], alert about vital signs being outside of a reference interval [11], support in dietary questions [5; 20], and many other critical issues. The effect of CDSSs on laboratory test ordering was assessed in several systematic reviews [6; 13; 15]. These reviews determined that, when integrated into clinical workflow, CDSSs have a positive influence on the process of test ordering by the doctors. However, when the decision to run a diagnostic test or to follow up on a test is switched from a doctor to a patient, the effect of a CDSS can be very strong. Direct access of patients to the test results can lead to the patients becoming better-informed [4], more engaged [8], and able to manage their care more efficiently [21]. Another potential benefit is improving patient safety. Casalino et al. discovered that 8–26% of abnormal test results are not followed up in a timely way

¹ Corresponding Author, Georgy Kopanitsa, Birzhevaya liniya 4, Saint-Petersburg, Russia, georgy.kopanitsa@gmail.com.

[3]. Pillemer et al. demonstrated that access to test results is appreciated by patients and increases adherence. However, it may cause patient nervousness and increase the number of visits [12]. Direct access to the test findings may let patients pursue a proper follow-up rate. A major concern expressed in [18] is that direct access can cause excessive patient nervousness because they can face difficulties interpreting the results. Interpretations given by a CDSS can support patients as they proved to deliver correct interpretations, provide great user experience, and are generally accepted by the patients [16]. A good and timely provided interpretation of the laboratory tests results given to patients by a CDSS, especially in the situation when a patient does not visit a doctor after the test is done can help to decide how to proceed with the diagnosis without support from a healthcare professional. The objective of the research is to study how patients perceive the interpretations of the laboratory tests automatically generated by a clinical decision support system in comparison with the interpretations given by doctors and how this affects the follow-up rate.

2. Methods

A clinical decision support system provides interpretation of laboratory test results for patients [11]. The common regional patient registry with obligatory registration of all laboratory tests helps to identify all testing events for every patient, including a list of tests that the patient ordered. We conducted a retrospective medical record review of 3200 randomly selected patients aged 18 to 76 querying the common regional patient registry. One patient could be included (1 clinical case) in the study only once. We studied 4 groups of patients:

1. The patients from the first group received their test interpretations automatically generated by a clinical decision support system with a clear indication that the interpretation was done automatically by a clinical decision support system.
2. The patients from the second group received their interpretations personally from a doctor with a clear indication that the interpretation was done automatically by a clinical decision support system.
3. The patients from the third group received their interpretations from a doctor with no indication of their automated generation.
4. The patients who did not receive any recommendations, only the test results.

We selected six laboratory tests commonly done by the patients (86.4% of all laboratory tests ordered in 2018), for which a decision support system could generate interpretations. Interpretation of each test with abnormal results included a recommendation to run another specific laboratory test or to repeat the same test again depending on the conditions of the patient. Participants were randomly selected for each group, using Random Sampling method, that ensured comparability to Russian demographic distributions. The reference intervals for the patients were taken from the recommendations of the Ministry of Health of the Russian Federation. We considered test results to be abnormal only if they were at least 20% outside of the reference range because it indicated a danger or potential risk for the patient's well-being over time.

A follow-up rate was calculated as a ratio of the number of patients who referred to a laboratory service for a follow-up investigation after receiving a recommendation within two weeks after the first test with abnormal test results had been completed and the interpretation was delivered to the patient. We calculated only the follow-up rate for the patients that had a recommendation to run additional laboratory tests. We invited patients

from groups 1 and 2 to take part in interviews to gain a deeper understanding of the motives to follow up or not after receiving a recommendation and the role of the automatically generated interpretation in this decision. The patients were invited by email and personal telephone calls. Patients were eligible for the study if they were 18 years and older and received a laboratory test interpretation during the previous months. We invited all the patients (1600) from groups 1 and 2 to participate in the interviews, 836 of them accepted the invitations: 421 from group 1 and 415 from group 2. From the patients who agreed, we formed a study group to represent the initial study population in terms of gender and age. Based on these constraints, we formed a group of 689 patients to run the study. The study was performed using one-on-one structured interviews. We asked the patients to rate questions 2–5 based on a Likert scale from 1 (not at all) to 7 (very much) [7]. For question 1, there were two possible answers: “yes” and “no”. The interviews were done by one researcher by phone or in person. The interview agenda was developed by the project team and reviewed and accepted by the ethics committee of the regional Department of Health. Before the interview, the patients were informed about the goals of the research and received a printed declaration of anonymity and confidentiality. Every patient had signed a consent form before the interviews began. Interview agenda for the patients consisted of the following questions:

- 1. Q1: Did you notice that the recommendation was generated automatically?
- 2. Q2: Do you think the fact that the interpretations are done automatically affect your decision to follow up?
- 3. Q3: Do you think doctors will give a more precise and valid interpretation?
- 4. Q4: Would you wait to get a recommendation from a doctor rather than get an immediate interpretation from a CDSS?
- 5. Q4: Do you trust the interpretation given be a clinical decision support system?

3. Results

The demographic data of the patients were recorded from the regional patient registry and presented in Table 1.

Table 1. Demographic characteristics of the patients

Gender	Group 1		Group 2		Group 3		Group 4		Age	Total
Age	>=60	<60	>=60	<60	>=60	<60	>=60	>60		
	19.4%	80.6%	19.4%	80.6%	19.4%	80.6%	19.4%	80.6%		
Males	47.0%		46.75%		47.38%		46.38%		38.8	46.87%
	74	302	70	304	76	303	72	299		1500
Females	53.0%		53.25%		53.62%		53.62%		39.4	55.13%
	97	327	95	331	97	324	99	330		1700
Total	171	629	165	635	173	627	171	629		3200

The follow-up rates for every group of patients are shown in Table 2.

Table 2. Laboratory test follow-up rate for every group

Gender		Group 1		Group 2		Group 3		Group 4	
	Age	>60	<60	>60	<60	>60	<60	>60	<60
Males		59%	60%	73%	74%	75%	79%	48%	42%
Females		61%	61%	72%	75%	73%	80%	55%	49%
Total, age-dependent		55%	61%	72%	75%	74%	80%	49.7%	46%
Total		58%		74%		78%		49%	

Average follow-up 68%
with interpretations

Patients’ attitude to the automatic test interpretations is presented in the table 3.

Table 3. Patients’ attitude to the automatic test interpretations

Question	Group 1		Group 2		Total	
	Mean	Median	Mean	Median	Mean	Median
Q1	3.4	4	3.1	3	3.3	3
Q2	6.8	7	6.1	7	6.2	7
Q3	3.1	4	3.4	3	3.3	3
Q4	5.9	5	5.7	5	6.8	5

4. Discussion and conclusion

Patients are generally willing to have access to their test results [16], so the automatically generated interpretation, especially in the case when a patient is not visiting a doctor, can enhance their experience and increase follow-up rates. Callen et al. in their systematic review revealed that 6.8–62% of laboratory tests were not followed up by patients [2; 17]. Utilization of clinical decision support systems for test interpretations can assist in overcoming the problem of failures to inform patients of abnormal outpatient findings, which is a common case now according to Casalino et al. with up to 26.2% of abnormal results not delivered to patients [3]. Automatic delivery of test result interpretations can potentially decrease the number of uninformed patients. The results of our research demonstrate that there is a significant impact on the patients’ decision to follow up on the tests. Promotion of laboratory test result notifications with detailed interpretations still demands communication effort and improvement of the algorithms to increase the reliability and trust to the patient-facing clinical decision support systems.

5. Acknowledgements

This work financially supported by the government of the Russian Federation through the ITMO fellowship and professorship program. This worked was supported by a Russian Fund for Basic research 18-37-20002.

References

[1] S. Breen, D. Ritchie, P. Schofield, Y.S. Hsueh, K. Gough, N. Santamaria, R. Kamateros, R. Maguire, N. Kearney, and S. Aranda, The Patient Remote Intervention and Symptom Management System (PRISMS) - a Telehealth- mediated intervention enabling real-time monitoring of chemotherapy side-effects in patients with haematological malignancies: study protocol for a randomised controlled trial, *Trials* **16** (2015), 472.

[2] J.L. Callen, J.I. Westbrook, A. Georgiou, and J. Li, Failure to follow-up test results for ambulatory patients: a systematic review, *J Gen Intern Med* **27** (2012), 1334-1348.

- [3] L.P. Casalino, D. Dunham, M.H. Chin, R. Bielang, E.O. Kistner, T.G. Karrison, M.K. Ong, U. Sarkar, M.A. McLaughlin, and D.O. Meltzer, Frequency of failure to inform patients of clinically significant outpatient test results, *Arch Intern Med* **169** (2009), 1123-1129.
- [4] L. Covolo, S. Rubinelli, E. Ceretti, and U. Gelatti, Internet-Based Direct-to-Consumer Genetic Testing: A Systematic Review, *J Med Internet Res* **17** (2015), e279.
- [5] M.E. Cupples, J.A. Cole, N.D. Hart, N. Heron, M.C. McKinley, and M.A. Tully, Shared decision-making (SHARE-D) for healthy behaviour change: a feasibility study in general practice, *BJGP Open* **2** (2018), bjgpopen18X101517.
- [6] N. Delvaux, K. Van Thienen, A. Heselmans, S.V. de Velde, D. Ramaekers, and B. Aertgeerts, The Effects of Computerized Clinical Decision Support Systems on Laboratory Test Ordering: A Systematic Review, *Arch Pathol Lab Med* **141** (2017), 585-595.
- [7] B.L. Drinkwater, A comparison of the direction-of-perception technique with the Likert method in the measurement of attitudes, *J Soc Psychol* **67** (1965), 189-196.
- [8] E.G. Feil, P.G. Sprengelmeyer, B. Davis, and P. Chamberlain, Development and testing of a multimedia internet-based system for fidelity and monitoring of multidimensional treatment foster care, *J Med Internet Res* **14** (2012), e139.
- [9] M. Finnerty, E. Austin, Q. Chen, D. Layman, E. Kealey, D. Ng-Mak, K. Rajagopalan, and K. Hoagwood, Implementation and Use of a Client-Facing Web-Based Shared Decision-Making System (MyCHOIS-CommonGround) in Two Specialty Mental Health Clinics, *Community Ment Health J* **55** (2019), 641-650.
- [10] M. Ghodusi Johari, M.H. Dabaghmanesh, H. Zare, A.R. Safaiean, and G. Abdollahifard, Smart Diabetic Screening and Managing Software, A Novel Decision Support System, *J Biomed Phys Eng* **8** (2018), 289-304.
- [11] G. Kopanitsa and I. Semenov, Patient facing decision support system for interpretation of laboratory test results, *BMC Med Inform Decis Mak* **18** (2018), 68.
- [12] F. Pillemer, R.A. Price, S. Paone, G.D. Martich, S. Albert, L. Haidari, G. Updike, R. Rudin, D. Liu, and A. Mehrotra, Direct Release of Test Results to Patients Increases Patient Engagement and Utilization of Care, *PLoS One* **11** (2016), e0154743.
- [13] P.S. Roshanov, J.J. You, J. Dhaliwal, D. Koff, J.A. Mackay, L. Weise-Kelly, T. Navarro, N.L. Wilczynski, R.B. Haynes, and C.S.R. Team, Can computerized clinical decision support systems improve practitioners' diagnostic test ordering behavior? A decision-maker-researcher partnership systematic review, *Implement Sci* **6** (2011), 88.
- [14] M.A. Rubin, K. Bateman, S. Donnelly, G.J. Stoddard, K. Stevenson, R.M. Gardner, and M.H. Samore, Use of a personal digital assistant for managing antibiotic prescribing for outpatient respiratory tract infections in rural communities, *J Am Med Inform Assoc* **13** (2006), 627-634.
- [15] P.J. Scott, A.W. Brown, T. Adediji, J.C. Wyatt, A. Georgiou, E.L. Eisenstein, and C.P. Friedman, A review of measurement practice in studies of clinical decision support systems 1998-2017, *J Am Med Inform Assoc* (2019).
- [16] I. Semenov, G. Kopanitsa, D. Denisov, Y. Alexandr, R. Osenev, and Y. Andreychuk, Patients Decision Aid System Based on FHIR Profiles, *J Med Syst* **42** (2018), 166.
- [17] I. Semenov, G. Kopanitsa, A. Karpov, G. Lakovenko, and A. Laskovenko, Implementation of a Clinical Decision Support System for Interpretation of Laboratory Tests for Patients, *Stud Health Technol Inform* **224** (2016), 184-188.
- [18] S. Sung, V. Forman-Hoffman, M.C. Wilson, and P. Cram, Direct reporting of laboratory test results to patients by mail to enhance patient safety, *J Gen Intern Med* **21** (2006), 1075-1078.
- [19] B. Tucker Edmonds, S.M. Hoffman, D. Lynch, E. Jeffries, K. Jenkins, S. Wiehe, N. Bauer, and M. Kuppermann, Creation of a Decision Support Tool for Expectant Parents Facing Threatened Periviable Delivery: Application of a User-Centered Design Approach, *Patient* **12** (2019), 327-337.
- [20] M.M. Warner, J.T. Kelly, D.P. Reidlinger, T.C. Hoffmann, and K.L. Campbell, Reporting of Telehealth-Delivered Dietary Intervention Trials in Chronic Disease: Systematic Review, *J Med Internet Res* **19** (2017), e410.
- [21] A. Woywodt, K. Vythelingum, S. Rayner, J. Anderton, and A. Ahmed, Single-centre experience with Renal PatientView, a web-based system that provides patients with access to their laboratory results, *J Nephrol* **27** (2014), 521-527.

Clinical-Based and Expert Selection of Terms Related to Depression for Twitter Streaming and Language Analysis

Angela LEIS^a, Miguel-Angel MAYER^a, Francesco RONZANO^a, Marta TORRENS^{b,c},
Claudio CASTILLO^b, Laura I. FURLONG^a and Ferran SANZ^{a,1}

^a *Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM) and Universitat Pompeu Fabra, Barcelona, Spain*

^b *Institute of Neuropsychiatry and Addictions, IMIM, Barcelona, Spain*

^c *Psychiatric Department of Universitat Autònoma de Barcelona, Spain*

Abstract. People use language to express their thoughts and feelings, unveiling important aspects of their psychological traits and social interactions. Although there are several studies describing methodologies to create a collection of words in English related to depression and other conditions, in most of them the selection of words is not clinical or expert based. The objective of this study is twofold: firstly, to introduce a comprehensive collection of Spanish words commonly used by patients suffering from depression, which will be available as a free open source for research purposes (GitHub), and secondly, to study the usefulness of this collection of words in identifying social media posts that could be indicative of patients suffering from depression. The level of agreement among medical doctors to determine the best words that should be used to select tweets related to depression was low. This finding may be due to the complexity of depression and the extraordinary diversity in the way people express themselves when describing their illness. It is critical to perform a thorough analysis of the specific language used in each condition, before deciding the best words to be used for filtering the tweets in each disease. As our study shows, the words supposedly more linked to depression are very common words used in other contexts, and consequently less specific for detecting depressive users. In addition, grammatical gender forms should be considered when analysing some languages such as Spanish.

Keywords. Depression, social media, surveys and questionnaires, terminology

1. Introduction

People use words to reflect their thoughts and feelings, revealing a huge amount of information about their personality and social interactions, as well as different psychological traits [1]. Language is the medium by which mental health professionals attempt to understand human beings and mental disorders. Several studies have found that linguistic styles are indicative of depressed mood [2]. Another interesting point is that people's language is stable over time and consistent across subjects or context; consequently, it can be used as a tool to measure differences among individuals [3].

¹ Corresponding Author: E-mail: ferran.sanz@upf.edu.

Several studies used social media platforms such as Twitter to analyze psychiatric symptoms and diseases, including depression [4,5]. Although these studies describe different methodologies to create a collection of words in English related to depression and other mental disorders, in the majority of them the selection of words were not extensive or clinical and expert-based [6,7]. In addition, Spanish speaking countries, such as Spain and Mexico, are among the ten countries with most Twitter users worldwide, with more than 6 million and 7 million users, respectively [8]. As far as we know, there are no studies focused on the creation of a list of words of depression in Spanish. The selection of the best terminology and keywords in any information system, including electronic health records, bibliographic databases or social media platforms, is critical for their usefulness in management and scientific research.

The objective of this study is twofold: first, to create a comprehensive collection of Spanish words commonly used by patients suffering from depression, which will be used for streaming Twitter and that is available as a free open source for research purposes in GitHub. Second, we determine the usefulness of the different words in identifying social media posts potentially related to depression.

2. Methods

The methodological approach consisted of three phases. The first phase consisted of the review of the most common tests and rating scales for assessing depression (using their Spanish version) in order to extract terms related to the expression of the depression symptoms. The different tests and rating scales analyzed are shown in table 1. The review was carried out by a psychologist and a family physician with experience in the clinical characteristics of depression, who reached an agreement on the words representative of the language used by depressive patients. The Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) [9] was reviewed to complete the list. In this way, a list of 255 words was created.

Table 1. Tests and rating scales used for selecting words related to the expression of depression by patients (the Spanish versions were used)

Name of the tests and rating scales
Beck Depression Inventory (BDI)
Brief Symptom Inventory (BSI)
Carroll Rating Scale for Depression
Center for Epidemiologic Studies Depression Scale (CESD-R)
Clinically Useful Depression Outcome Scale (CUDOS)
Goldberg Depression and Anxiety scales (GADS)
Hamilton Rating Scale for Depression (HRSD)
Hospital Anxiety and Depression de Zigmond and Snaith (HAD)
Montgomery-Asberg depression rating scales (MADRS)
The Patient Health Questionnaire (PHQ-9)
Zung Self-Rating Depression Scale (SDS)

In the second phase, a questionnaire was created including the aforementioned list of words. The questionnaire was sent in December 2017 via email to 50 psychiatrists from the Institute of Neuropsychiatry and Addiction (INAD) of Parc Salut Mar in Barcelona and 5 family physicians from the Spanish National Health Service. The email included the purpose of the questionnaire, the guidelines for its completion and an Excel file with the words selected. A second email was sent as a reminder in February 2018. The objective of this questionnaire was to obtain a score that represented how well these

words are related to depression, as described by patients explaining their symptoms in clinical settings. The score of a word was obtained by adding the scores provided by each rater using a Likert scale (1. never, 2. rarely, 3. occasionally, 4. frequently and 5. very frequently). Since 20 raters participated in the scoring process, the maximum value of the score was 100 points. The list of 255 words and their scores are available at GitHub for research purposes: <https://github.com/angelaleism/WordsDepression>.

Finally, the third phase involved streaming of tweets that included at least one of the 375 words (the list of 255 words plus their different plural and gender forms), which was carried using the Twitter Application Programming Interface (API) [10]. The streaming was set up between June and September 2018 obtaining 8,832,256 tweets. In order to compare the usefulness of the words to detect tweets with signs of depression, two sets of 500 tweets were randomly selected after retweets removal. One of the sets was created by including five subsets of 100 tweets, each one composed by tweets that included one of the five most highly scored words, and the second 500-tweet set included low scored words. The 1,000 tweets were manually reviewed by two experts to determine whether they were potentially indicative of depression or not. The R programming language was used for the statistical analyses.

3. Results

The questionnaire was answered by 30% of psychiatrists (15/50) and all the family physicians (5/5; 100%). The respondents were 13 women and 7 men. All the respondents rated the complete list of 255 words. The mean and SD of the scores of all the participants are shown in figure 1.

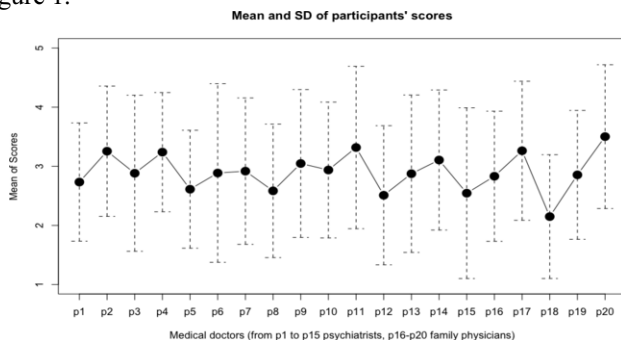


Figure 1. Distribution of the means and standard deviations (SD) of the 20 participants' scores

In order to assess the reliability of agreement among the raters, we calculated the intraclass correlation coefficients (ICCs) and the agreement among all the raters was 0.47, for the psychiatrists was 0.53 and for the family physicians was 0.37.

Based on the health professionals' scores, the 10 words in Spanish most frequently expressed by depressive patients were (the translation in English is in parenthesis): *deprimido/a* (depressed), *triste* (sad), *tristeza* (sadness), *desanimado/a* (downhearted), *depression* (depression), *depresivo/a* (depressive), *ansiedad* (anxiety), *cansado/a* (tired), *lloro* (crying), *insomnio* (insomnia). The 10 less frequent words were: *autocrítico* (self-criticism), *ingrato/a* (ungrateful), *miserio/a* (vile), *languidez* (languid), *mutilado/a* (disabled), *apetencia* (hunger), *sombrio/a* (gloomy), *achacoso/a* (sickly), *desdeñado/a* (disdained), *lasitud* (weariness). It is necessary to take into account that in the translation

of these words into English there are some nuances that may be missing. There are some words that are more frequent in its feminine form on Twitter such as anxious, distrustful, distressed, insecure or shy and more frequent in masculine such as loser, solitary, incompetent or defeated.

Regarding the analysis of the tweets, table 2 shows the number of tweets that included the studied words, as well as the scores assigned to these words by the health professionals and the ranking of the words on the basis of the scores. In addition, the table shows the proportion of tweets for each word that were potentially indicative of depression when manually reviewed by an expert.

Table 2. Frequencies of words in the 8,832,256 tweets analysed, scores and proportion of depressive tweets

Word	Number of tweets	Score (Rank)	% depressive tweets
Deprimido/a	63,019	97 (1)	12%
Triste	56,776	97 (2)	5%
Tristeza	54,789	96 (3)	10%
Desanimado/a	24,079	94 (4)	23%
Depresión	51,408	92 (5)	12%
Infelicidad	12,116	63 (110)	4%
Suicida	42,615	57 (127)	1%
Vencido/a	43,020	52 (162)	3%
Melancolía	22,896	43 (194)	9%
Desdichado/a	4,663	41 (203)	9%

4. Discussion and conclusions

The diagnosis of depression is a complex process because of the heterogeneous nature of this disease, the lack of biological markers, the different symptoms among individuals and the diverse ways in which patients express those symptoms.

In relation with the scores assigned by the health professionals that participated in the survey, there was more agreement among the psychiatrists than among the family physicians. These results may be consistent with the fact that the psychiatrists deal with more patients with depressive disorders, and therefore, they are more familiar with the language used by patients with depression. As a result, the agreement between psychiatrists and family doctors is low. This finding may be due to the extraordinary diversity in the way people express themselves when describing their illness, the complexity of depression [9], how health professionals interpret these words for making a diagnosis and their clinical experience. For this reason, it is critical to perform a thorough analysis of the specific language used in each condition, before deciding the most suitable words to be used for filtering the tweets in each disease [2]. As our study shows, the words supposedly more linked to depression are very common words used in other contexts on Twitter, and consequently less specific for detecting depressive people (e.g. *triste/sad*). On the other hand, words less frequently mentioned on Twitter can have more weight to link them to depression or suicide tendency (e.g. *desanimado/downhearted*). The gender of words should be always considered when analyzing some languages that have grammatical gender forms such as Spanish.

Nevertheless, the analysis of words related to depression also require us to consider other aspects such as the linguistic features (i.e. the different use of personal pronouns, the number of negative words and the expressions associated to the basic emotions) and the behavioral patterns (i.e. the distribution of tweets over time and the number of characters or hashtags per tweet) [5]. However, the extensive list of words provided in

this study can be used as a basis for developing new studies and strategies for the analysis of depression on Twitter in Spanish. Patients can be monitored, introducing new opportunities for studying depression and providing additional health services.

This study presents some limitations. On the one hand, the tweets can be considered an indirect and inaccurate way of detecting users suffering from depressive disorders and it is not possible to verify whether the diagnosis is genuine or not. On the other hand, although the list of words was carefully developed, there may be more expressions or words not included in the list.

Acknowledgements

The Research Programme on Biomedical Informatics is a member of the Spanish National Bioinformatics Institute, PRB2-ISCIII, and it is supported by grant PT17/0009/0014. Special thanks to the psychiatrists of the Institute of Neuropsychiatry and Addiction (INAD) of Parc Salut Mar in Barcelona and the family community physicians who participated in the study.

References

- [1] C. Chung, J. Pennebaker. The Psychological Functions of Function Words. *Social Communication*, (2007), 343–359. Available at: <http://doi.org/10.4324/9780203837702>
- [2] T. Nguyen, M. Larsen, D. Phung, S. Venkatesh, H. Christensen. Using linguistic and topic analysis to classify sub-groups of online depression communities. *Multimedia Tools and Applications* **76**, (2017), 10653–10676. Available at: Available from: <http://doi.org/10.1007/s11042-015-3128-x>
- [3] W. Bucci, N. Freedman. The language of depression. *Bulletin of the Menninger Clinic* **45**(4), (1981), 334–358.
- [4] M. De Choudhury, M. Gamon, S. Counts, E. Horvitz. Predicting Depression via Social Media. In: *Proceedings of the Seventh International Conference on Weblogs and Social Media, AAAI'13*; July 8–11, 2013; Cambridge, MA p. 128–138.
- [5] A. Leis, F. Ronzano, M.A. Mayer, L.I. Furlong, F. Sanz. Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis. *Journal of Medical Internet Research* **21**(6), (2019), e14199.
- [6] V.M. Prieto, S. Matos, M. Álvarez, F. Cacheda, J.L. Oliveira. Twitter: a good place to detect health conditions. *PLoS One* **9**(1), (2014), e86191.
- [7] P.A. Cavazos-Rehg, M.J. Krauss, S. Sowles, S. Connolly, C. Rosas, M. Bharadwaj, et al. A content analysis of depression-related Tweets. *Computers in Human Behavior* **54**, (2016), 351–357
- [8] Statista. Leading Countries Based on Number of Twitter Users as of January 2019 (in millions) (2019). Available at: <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>
- [9] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Washington, DC: American Psychiatric Publishing; 2013.
- [10] Twitter Developer. Available at: <https://developer.twitter.com/en.html>

Design of the CAPABLE Health Empowerment Tool: Citizens' Needs and Expectations

Astri Letnes JANSON ^{a,1}, Anne MOEN ^b and Kristin Skeide FUGLERUD ^c

^a *Division of Diagnostics and Technology, Akershus University Hospital, Norway*

^b *Institute of Health and Society, University of Oslo, Norway*

^c *Department of Applied Research in Information Technology, Norwegian Computing Center, Norway*

Abstract. It is important but difficult for citizens to gain a proper overview of their health-related information. The CAPABLE project aims to create a digital tool that enable citizens to actively utilize their clinical and personal health information to manage medication, improve nutrition, and facilitate health services coordination. 57 participants have participated in various user-centered co-creation activities. A key finding of citizen's needs and expectations is that citizens are positive about such a tool provided it is easy to use and requires minimal work. Especially, opportunities to collect and utilize already existing digital health information, having access to personalized and relevant checklists and personalized resources, and be able to share health information with health personnel and others according to their own trust and privacy preferences. Involving other stakeholders to ensure integration with other systems is of principal importance to fulfil citizens' expectations of integration and automatization.

Keywords. User-centered design, inclusive design, mHealth, automatization, personalization, privacy, medication management, interoperability

1. Introduction

Today, citizens can access subsets of their health information through dedicated, secure patient portals [1], but content can be fragmented and access unequal [2], or requiring users to manually enter information into the apps of their choice [3]. Apps are usually targeting one specific condition, e.g., diabetes [4], heart failure [5] or cystic fibrosis [6] to mention a few. Over time, most people experience many health issues leading to health encounters, many of which produce information and knowledge that may be of relevance and importance for self-management and the best possible follow-up of treatments. The result is that citizens ends up managing a mix of paper documents, digital snippets of information exchanges or data in incompatible formats, supplemented with ad-hoc strategies for interpretation and management. The extent of health information exchange varies considerably, and is only partly achieved in some countries [7]. For citizens, there is currently no simple way to collect, curate or complement nor control health information from multiple institutions, services and systems, making it quite challenging to retain a good overview and understanding of its implications.

¹ Corresponding Author, Astri Letnes Janson, E-mail: astri.letnes.janson@ahus.no

We anticipate that easy access to and overview of all health-related information will contribute to increased empowerment and improved digital health literacy, which in turn can contribute positively to health status [6]. In our ongoing project, CAPABLE, we aim to support citizens' ability to manage their health and participate in health-related decisions. We are creating a mobile tool to enable citizens to actively utilize their clinical and personal health information to manage medication, improve nutrition, and facilitate health services coordination. To counteract current inequality in health services provision [8], we emphasize and pay special attention to the needs of vulnerable groups, people with disabilities, complex and or chronic conditions, and those with low health literacy. The research focus for the user-centered design process we report from was: what are the citizens' needs, values and the social context for developing the CAPABLE tool? What are the expectations of citizens for such a tool?

2. Method

The development of the CAPABLE tool is based on an inclusive user-centered design approach [9], emphasizing user involvement from the very beginning and throughout the full development and deployment cycle. Involving users in the design process contributes to increased applicability, accessibility, acceptance and adoption of the designed solution. It is considered especially efficient and influential in the early stages of the development process since the costs of making changes increases as the development continues [10]. We follow an interpretive research approach and utilize different user-centered methods. The variety and flexibility of methods allowed us to further investigate certain issues and questions, helping to clarify and identify recurring themes and topics. 57 citizens have participated in two focus groups (n=13), five personas workshops (n=15), a design workshop with five groups (n=24), a pluralistic usability walkthrough (n=4), and a paper prototype user testing (n=1). Municipalities affiliated with the project (n=3) recruited participants through existing municipal councils (elderly, disability, youth), and non- governmental disability and health organizations (n=3) recruited via their members. Participants represent citizens, from adolescence to elderly, with and without disabilities, with varying degrees of health problems, digital health literacy, and in different stages of life: pupils, workers, next of kin and retirees. We conducted a thematic analysis based on detailed field notes from all the activities.

3. Results

The analysis showed a wide range of expectations toward the CAPABLE tool – both affirming and contradictory to each other, ranging in level of detail and developmental implications. An overall, very common expectation was for the CAPABLE tool to require only small amounts of work to utilize it since many citizens experienced that creating an overview of their health information could be labor-intensive.

Providing citizens with a tool that requires them to register all their health information comes with little value, according to the participants in our research. They expected to be able to utilize the information that they are aware of already exists digitally. Presupposing imported digital health information, citizens expressed a willingness and eagerness to log additional, relevant health parameters related to their health challenges to gain knowledge and overview. In general, citizens are well aware

of the national summary care record “Kjernejournal” containing information from the last 3 years [1]. When they see themselves as healthy or have few or straightforward health issues, they find information about themselves in “Kjernejournal” as sufficient, but if conditions are more complex, they told that information could be incomplete or missing, requiring them to recall incidents from years or decades ago when asked.

Citizens anticipated that the CAPABLE tool could help them to manage experiences, questions and information before and after consultations with health personnel. To keep documentation to a minimum, citizens anticipate that checklists relevant to diagnosis, conditions or medicines could be useful. They pointed to examples of resources and information from the public health authorities or NGOs. However, CAPABLE tool should facilitate modifications in the resources it to their specific needs, according to their experiences and knowledge.

To support their memory in relevant situations, our participants would like to store health information in a secure place, where they can access it when they need to. At the same time, participating citizens were conscious of how their personal health data needs to be accessible for them, but not for their community. If functionality exists, they want to share their information with people of their own choice. The attractiveness of sharing relevant information, and decide how much information to share, for which treatment purposes and period, was high and seen as a feature for personal control.

4. Discussion

Citizens’ expect to utilize their health information with as much automation as possible comes with three main implications for design, discussed in the following.

4.1. Collecting Already Existing Digital Information

Our findings confirm previous research, which consistently find that ease of use, usability, accessibility and readability are key factors for successful adoption and use of eHealth systems [3,11-13]. Our participants go even further by emphasizing the need for and attractiveness of integration and utilization of existing digital health information. To collect existing digital health data, as we propose for the CAPABLE tool, the citizen could ask for access to personal health information from electronic sources, and save a copy of their health information for personal use, annotation, complementation, or even sharing personal health data at the person’s discretion. This approach complements established communication lines and should maintain data integrity in the source. CAPABLE build extensively on HL7 FHIR[®] resources to foster interoperability, the international leading interoperability standard, also recommended by the Norwegian Directorate for eHealth [14]. Above all, the CAPABLE tool will give the citizens an opportunity to maintain personal health information in one system, which was also a principle finding in Floch et al. [6].

4.2. Curating Personal Health Information

Personalization can contribute to making the application relevant and reduce the amount of work for each user. Kwon & Kim [15] differentiate between two types of personalization: *implicit*, based on previously collected consumer data, or *explicit* where

the consumer proactively specifies elements. In the CAPABLE tool, using an implicit approach could be possible as a starting, focal point of reference. Assumptions of health conditions can be delineated from a medication classification system, e.g. if a citizen take Carbamazepine, epilepsy is quite likely. However, the citizen might use the same drug for neuropathic pain or bipolar disorder. Moreover, the citizen may need information for conditions that cannot be deduced from the medication list. Therefore, explicit personalization would be a better strategy. As discussed by Li [16], incorrect information makes people more likely to interpret content as non-personalized, meaning that efforts to make content personalized in the CAPABLE tool can be futile. To balance these trade-offs, we propose a procedure where the citizens explicitly are asked to confirm deductions about health conditions. Further, we propose functionality where the citizens can select between predefined checklists and texts, and then add text alternatives that they miss to curate or adjust information to be meaningful for them.

4.3. Controlling Health Information

“Privacy by design” is important to comply with a legal perspective, but also paramount to the citizens’ maintained trust and willingness to use the CAPABLE tool. Recent research has found that many mHealth apps run the risk of jeopardizing the privacy of its users [17] because they do not conform to the General Data Protection Regulation (GDPR). CAPABLE needs to provide citizens with an easy and flexible way to control and share their information, while at the same time avoid unintended and unforeseen incidents. For sharing information from CAPABLE with health personnel, we plan to use SMART on FHIR to integrate CAPABLE with other health information systems if the citizen chooses to do so. This is well-aligned with The Norwegian Directorate of eHealth advice for our setting – that SMART on FHIR based on HL7 profiles is the most promising framework for application integration in EHR’s [14].

4.4. Limitations

Although this research has involved a relatively large number of participants, the current emphasis has been on the users of health services and their representatives. There is a need for further work to integrate this with the needs and expectations of other types of stakeholders. In addition, the views reported here reflect those who were willing to take part in this research project. People who were not interested or not able to participate may potentially have different viewpoints. A strength of the study is the combination of methods and the involvement of a wide diversity of citizens.

5. Conclusion

User involvement from citizen stakeholder assessment shows that it is of paramount importance to create and offer novel functionality going beyond “yet another registration tool”. Providing novel opportunities to collect and actively use digital personal health information to minimize text entry, be able to curate and annotate to meet personal needs and relevance, and ensure interoperability in privacy preserving manners could foster innovation that really empower the citizen. This user centered design approach has

informed the first step in designing the CAPABLE prototype. Based on the experiences, we will create a first prototype, piloting with approximately 50 additional users. Finally, we will carry out a large-scale feasibility trial with a proof of concept version of CAPABLE. The main challenge ahead is involving other stakeholders to meet the needs and expectations articulated from citizens.

Acknowledgements We acknowledge participants and project partners for their insights and contributions and the Research Council of Norway for funding.

References

- [1] The Norwegian Directorate of eHealth. Summary care record for safer healthcare in Norway; 2012 [cited 2019 sept 25]. Available from: <https://helsenorge.no/kjernejournal/kjernejournal-for-safer-healthcare>
- [2] A. Essén, I. Scandurra, R. Gerrits, G. Humphrey, M.A. Johansen, et al., Patient access to electronic health records: Differences across ten countries. *Health Policy Technol.* (2018). 7(1):44-56.
- [3] M.N.K. Boulos, A.C. Brewer, C. Karimkhani, D.B. Buller, R.P. Dellavalle, Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online J Public Health Inform.* (2014). 5(3):229
- [4] O. El-Gayar, P. Timsina, N. Nawar, W. Eid, Mobile Applications for Diabetes Self-Management: Status and Potential. *J Diabetes Sci Technol.* (2013). 7(1):247-62.
- [5] R.M. Creber, M.S. Maurer, M. Reading, G. Hiraldo, K.T. Hickey, S. Iribarren. Review and Analysis of Existing Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self-Care Management Using the Mobile Application Rating Scale (MARS). *JMIR mHealth uHealth*, (2016). 4(2):470-82.
- [6] J. Floch, A. Zetti, L. Fricke, T. Weisser, L. Grut, T. Vilarinho. User Needs in the Development of a Health App Ecosystem for Self-Management of Cystic Fibrosis: User-Centered Development Approach. *JMIR mHealth uHealth*. (2018). 6(5):e.113
- [7] T.H. Payne, C. Lovis, C. Gutteridge, C. Pagliari, et al. Status of health information exchange: a comparison of six countries. *J Glob Health* (2019). 9(2):0204279
- [8] H.F. McClintock, J.E. Kurichi, F.K. Barg, F.K. Krueger, P.M. Colletti, et al.. Health care access and quality for persons with disability: Patient and provider recommendations. *Disabil Health J.* (2018). 11(3): 382-9.
- [9] K.S. Fuglerud, Inclusive design of ICT: the challenge of diversity [dissertation]. [Oslo] Faculty of Humanities: University of Oslo; (2014). 282 p.
- [10] S. Kujala, Effective user involvement in product development by improving the analysis of user needs. *Behav Inf Technol.* (2008) 27(6):457-73.
- [11] H.B.M. Mohammed, D. Ibrahim, N. Cavus. Mobile device based smart medication reminder for older people with disabilities. *Qual Quant.* (2018). 52(2):1329-42
- [12] M. Madsen, R. El Kaliouby, M. Eckhardt, M.E. Hoque, et al., Lessons from participatory design with adolescents on the autism spectrum. In: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems. 2009 Apr; Boston, MA, USA: ACM, 2009. p. 3835-40.
- [13] C. Sahin, Rules of engagement in mobile health: what does mobile health bring to research and theory? *Contemp Nurse*, (2018). 54(4-5):374-87.
- [14] The Norwegian Directorate of eHealth. Recommendation for use of SMART on FHIR; 2019 [cited 2019 sept 25]; <https://ehelse.no/standarder/ikke-standarder/anbefaling-om-bruk-av-smart-on-fhir>
- [15] K. Kwon, C. Kim, How to design personalization in a context of customer retention: Who personalizes what and to what extent? *Electron Commer Res Appl.* (2012). 11(2):101-16.
- [16] C. Li. When does web-based personalization really work? The distinction between actual personalization and perceived personalization. *Comput Human Behav.* (2016). 54:25-33.
- [17] A. Papageorgiou, M. Strigkos, E. Politou, E. Alepis, et al., Security and Privacy Analysis of Mobile Health Applications: The Alarming State of Practice. *IEEE Access.* (2018). 6:9390-403.

Designing a Patient-Centered Notification System into a PHR

Mariana SIMÓN¹, Giuliana COLUSSI^a, Janine SOMMER^a, Lucila BRUCHANSKI^a,
Fernando PLAZZOTTA^a

^a *Health Informatics Department, Hospital Italiano de Buenos Aires, Argentina.*

Abstract. Personal health records could help patients overcome the anxiety related to gaps on health related information. The objective of this research was to explore what type of information was perceived by patients as useful to receive as a notification on their PHR as a framework for the design of a patient-centered notification system. We applied a qualitative approach followed by usability tests. From interviews to patients, we obtained main categories of notifications to include according to its perceived usefulness while, from iterative usability tests, we obtained a final version of prototypes developed according to patient-interface interaction experiences. This research helped us to understand patients needs and behavior to design an efficient notification system and also represents the first step for the development of clinical decision support systems for patients into the PHR with the potential to improve clinical outcomes and healthcare delivery.

Keywords. Personal Health Records, Software Design, Patient preferences

1. Introduction

At present, patients are increasingly demanding information and effective communication in order to understand the background in which healthcare is taking place [1, 2]. In this context, personal health records (PHR) are tools developed to meet these needs [1,3], helping patients to overcome the anxiety related to gaps on health information [4]. Besides, PHR contributes to the interaction of patients and families with healthcare systems, facilitating a shared decision-making process and empowerment [5, 6].

As highlighted by Johnson et. al. [7], patients usually appreciate receiving test results notifications at the appropriate time, the availability of a direct communication channel with providers, and any tool that could facilitate healthcare management. Delays related to these needs might lead to increased anxiety generated by lack of information.

The objective of this research is to explore what type of information is perceived by patients as useful to receive as a notification or reminder on their PHR. The latter, as a framework for the design of a notification system applying user-centered design techniques, to ensure information availability at the appropriate time.

¹ Corresponding Author: E-mail: mariana.simon@hospitalitaliano.org.ar.

2. Methods

Hospital Italiano de Buenos Aires (HIBA) is an academic non-profit tertiary level organization that covers all health specialities at different levels of the care system. Since 1998, HIBA has his own health information system (in-house) [8]. An integrated PHR is available since 2007. Its main functionalities include: appointment scheduling, test results visualization, patient-physician secure messaging, management of prescribed medication, and different modalities of teleconsultation. At present, our PHR has approximately 400,000 registered users [9].

We conducted a mixed-methods research divided into 2 stages. On the first stage, we performed semi-structured individual interviews with patients to explore what type of information they found useful to receive as a notification or reminder into their PHR, as well as the preferred means to receive it. Once the data had been collected, we defined notifications categories determining the system functionalities that would support them and finally, we designed prototypes of the different interfaces.

At the second stage we applied a user-centered interaction approach to adjust prototypes design and usability. This approach gave as a clear input on how patients used the system. The interaction of users with prototypes was suitable to detect usability problems. Test participants were asked to read aloud 4 tasks related to notification functionalities and tried to solve them. At the same time, they were requested to describe the actions they were performing, expressing their thoughts as regards the interaction with interfaces. These tests included 3 participants: the user, the facilitator, and the observer.

Both, interviews and usability tests, were performed on a convenience sample of adult outpatients on different hospital waiting areas and dates. Both were recorded previous oral informed consent. The data was kept anonymous, protecting confidentiality of participants. In what concerns to data-analysis, we proceeded to analyze the previously recorded audios and categorize the information in domains according to the main findings. Data collection and data analysis were initiated simultaneously. The analytic process was performed through codification and categorization of the information, based on a process of constant comparison. The categories were defined taking into account both the research main questions and the findings that emerged from the interviews [10].

To analyze usability tests, we classified users' problems to fulfill the given tasks using a severity 3 point scale (critical, serious and minor), and we complemented this information with participants' suggestions.

3. Results

15 interviews were conducted on patients (table 1). 80% were female and age range was from 31 to 78 years old. 100% of participants were PHR users.

Table 1 - Patients characteristics

	n: 15
Female	12 (80%)
Age	
31-50	4 (26,7%)
51-70	9 (60%)
+70	2 (13,3%)
PHR users	15 (100%)

When participants were asked about those subjects/information they considered as useful to be notified of, 5 main categories emerged:

3.1 Appointment scheduling management

Patients explained that they found suitable to be notified about their closest medical appointments and any changes on them in advance. They considered this remainder will facilitate their own time management according to health visits.

3.2 Medication management

Interviewees considered as valuable to receive notifications whenever they need to request their physician for a prescription update before it expires. Besides they expressed the need of a tracking system, with notifications, for home delivery of medication.

3.3 Test results management

Participants mentioned as useful to receive notifications whenever test results were informed and available to read into the PHR. Also, they highlighted as an advantage, to be notified when a complementary study order was close to expire.

3.4 Administrative processes

Users expressed they would appreciate being notified on different administrative procedures needed to receive medical care, such as bill payment information and help desk answers upon their requests for support, among others.

3.5 Hospital Information

Hospital news and activities were other topic in which a notification process was found useful. Another type of information included institutional changes or system

downtimes. As regards the mean by which they preferred to be notified, we didn't find consensus in favor of a unique mean of notification. Some participants chose the email, while others preferred push notifications on their mobile PHR. Besides, they suggested that the mean of notification could vary according to the type of notification.

With the information collected from patients inquiries we defined notification categories. For the notification system design, we opted for a fully customizable feature, meaning each user can configure what notification to turn on or off, upon their preferences along with the preferred mean to receive them (email, push notification or mobile PHR).

Departing from designed prototypes, we conducted 3 iterations of usability tests with final users of different ages and technological literacy levels. Each iteration cycle resulted in prototypes modifications to overcome usability problems detected until reaching a final proposal. Each testing cycle was carried out with different users. 6 patients participated in the first set of usability tests and, its results, evidenced the presence of 3 degrees of usability problems: critical, serious and minors. After some adjustments that emerged from the analysis of these results, we performed a second iteration with the participation of 8 final users. Analysis of results showed that minor problems were still present. With this in mind, we completed another set of prototype adjustments that were tested, in a third iteration, with 8 patients. After this cycle no usability problems were detected and thus, we arrived to a final version of prototypes.

4. Discussion

Results obtained from this research allowed us to design a patient-centered notification system into our PHR. From interviews with patients, we obtained 5 main categories of notifications to include in the system according to its perceived usefulness. Besides, performing iterations of usability tests to patients, we obtained a final version of prototypes developed according to patient-interface interaction experiences.

Usability evaluation and user experience are key points for design and development of PHR functionalities [11,12,13]. In line with this fact, the results of our research, contributed to a better PHR user experience. Final versions represent an efficient design adjusted to user's needs.

The development of this functionality for the PHR, might enhance patient empowerment by displaying suitable information in time, optimizing the management of health-related aspects and decision making. A notification system gives patients the option to choose what to receive and when, acting according to the delivered information. PHRs could send notifications and reminders and contribute to a better healthcare without generating alerts fatigue [14]. In concordance with our findings as regards preferred functionalities to be notified, the implementation of a notification system for test results avoids important results to be overlooked [15].

We also arrived to similar conclusions concerning means of notification, that is, the preferred mean depends on the type of notification [15]. Furthermore, evidence suggests that message or notification systems to patients can improve outcomes such as infant vaccination coverage by sending reminders to patients [16] and increase smoking cessation rates [17]. This research also represents the first step for future research for the development of clinical decision support systems for patients into the

PHR. These tools might contribute to the shared-decision model with the potential to improve clinical outcomes and healthcare delivery.

References

- [1] Elder NC, Barney K. “But What Does It Mean for Me?” Primary Care Patients’ Communication Preferences for Test Results Notification. *The Joint Commission Journal on Quality and Patient Safety* (2012), 168–AP1.
- [2] Choudhry A, Hong J, Chong K, Jiang B, Hartman R, Chu E, et al. Patients’ Preferences for Biopsy Result Notification in an Era of Electronic Messaging Methods. *JAMA Dermatol.* **151** (2015), 513.
- [3] Tavares J, Oliveira T. New Integrated Model Approach to Understand the Factors That Drive Electronic Health Record Portal Adoption: Cross-Sectional National Survey. *J Med Internet Res.* **20** (2018), e11032.
- [4] Baldwin DM, Quintela J, Duclos C, Staton EW, Pace WD. Patient preferences for notification of normal laboratory test results: a report from the ASIPS Collaborative. *BMC Fam Pract.* **6** (2005), 11.
- [5] Johnson AJ, Easterling D, Nelson R, Chen MY, Frankel RM. Access to Radiologic Reports via a Patient Portal: Clinical Simulations to Investigate Patient Preferences. *Journal of the American College of Radiology.* (2012), 256–263.
- [6] Halamka JD, Mandl KD, Tang PC. Early Experiences with Personal Health Records. *Journal of the American Medical Informatics Association.* (2008), 1–7.
- [7] Stockwell MS, Fiks AG. Utilizing health information technology to improve vaccine communication and coverage. *Hum Vaccin Immunother* **9** (2013), 1802–1811.
- [8] Plazzotta F, Luna D, González Bernaldo de Quirós F. [Health information systems: integrating clinical data in different scenarios and users]. *Rev Peru Med Exp Salud Publica* **32** (2015), 343–351.
- [9] Smith MI, Garcia G, Simon M, Bruchanski L, Frangella J, Sommer JA, et al. Lessons Learned After Redesigning a Personal Health Record. *Stud Health Technol Inform.* **245** (2017), 216–220.
- [10] Sampieri R, Collado C, and Lucio P, *Metodología de la investigación*, McGraw-Hill Companies, 2006.
- [11] Silhavy P, Silhavy R, Prokopova Z. Patients’ Perspective of the Design of Provider-Patients Electronic Communication Services. *International Journal of Environmental Research and Public Health* (2014), 6231–6245.
- [12] Minoletti S, Rapisarda R, Giraldo L, Grande M, Sommer J, Plazzotta F, et al. User-Centered Design of a Pediatric Vaccination Module for Patients. *Stud Health Technol Inform.* **264** (2019), 1096–1100.
- [13] Zhou L, DeAlmeida D, Parmanto B. Applying a User-Centered Approach to Building a Mobile Personal Health Record App: Development and Usability Study. *JMIR Mhealth Uhealth* **7** (2019), e13194.
- [14] Hess R, Fischer GS, Sullivan SM, Dong X, Weimer M, Zeith C, et al. Patterns of response to patient-centered decision support through a personal health record. *Telemed J E Health* **20** (2014), 984–989.
- [15] Shultz SK, Wu R, Matelski JJ, Lu X, Cram P. Patient Preferences for Test Result Notification. *J Gen Intern Med.* **30** (2015), 1651–1656.
- [16] Domek GJ, Contreras-Roldan IL, Bull S, O’Leary ST, Bolaños Ventura GA, Bronsert M, et al. Text message reminders to improve infant immunization in Guatemala: A randomized clinical trial. *Vaccine* **37** (2019), 6192–6200.
- [17] Rodgers A. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. *Tobacco Control* (2005), 255–261.

Developing an Intervention to Implement Electronic Patient-Reported Outcomes in Renal Services in the UK

Sabine N VAN DER VEER^{a,1}, Angelo ERCIA^a, Fergus J CASKEY^b, Ken FARRINGTON^c, Francine JURY^a, Michael REES^d, Tim Whitlock^e and Sarah KNOWLES^f

^aCentre for Health Informatics, University of Manchester, UK

^bPopulation Health Sciences, Bristol Medical School, University of Bristol, UK

^cRenal Unit, East & North Hertfordshire NHS Trust, Stevenage, UK

^dSchool of Social, Historical and Political Studies, University of Wolverhampton, UK

^eUK Renal Registry, The Renal Association, Bristol, UK

^fHealth Sciences, University of York, UK

Abstract. Routinely collecting and using electronic patient-reported outcome (ePRO) data in clinical practice can improve patients' experience and outcomes, but implementing this at scale has proved challenging. As part of the *Optimising routine collection of electronic patient-reported outcomes* (OPT-ePRO) study, we therefore developed an intervention that aimed to facilitate the implementation of ePROs. We are conducting OPT-ePRO in the context of secondary care for people with chronic kidney disease in the UK, with three renal units participating as our study sites. Intervention design was guided by Normalisation Process Theory, and informed by published literature and qualitative research. The intervention consisted of a national infrastructure to securely collect, transfer and display ePRO data, complemented with materials and procedures to support kidney patients and renal unit staff with embedding ePROs in usual care pathways. The next step will be to bring the OPT-ePRO intervention into practice and iteratively refine it.

Keywords. eHealth; Patient-generated health data; Symptom assessment

1. Introduction

Electronic patient-reported outcome (ePRO) data is digitally collected information that reflects the personal impact of illness and treatment as assessed by patients, such as information on symptom burden or quality of life. Studies have shown that routinely collecting and using ePRO data as part of clinical care can improve patients' experiences and outcomes [1], while also informing audits and commissioning of services. The challenge now is to implement ePROs more widely in order to harness these potential benefits.

Despite extensive knowledge on how to successfully implement ePROs [2,3], national initiatives have struggled with low response rates [4,5]. One issue is that patient and staff engagement in the implementation varies widely between groups and

¹Corresponding Author, Sabine N van der Veer, E-mail: sabine.vanderveer@manchester.ac.uk

facilities, and is often limited [4,5]. This may be explained by insufficient support for embedding collection and use of ePROs into usual care pathways. Therefore, we developed an intervention aimed at facilitating implementation of ePROs into clinical practice, with UK renal services as the exemplar context. This formed the first phase of the *Optimising engagement of routine collection of electronic patient-reported outcomes* (OPT-ePRO).

2. Methods

2.1. Theoretical framework

We used Normalisation Process Theory (NPT) as the theoretical framework to guide development of the OPT-ePRO intervention [6]. NPT has been widely used to plan implementations of eHealth interventions in clinical practice [7] and consists of four constructs: coherence (e.g. do users have a shared view of the intervention's purpose); cognitive participation (e.g. do users understand and agree on who will deliver the intervention); collective action (e.g. do existing resources and systems allow the intervention to 'fit'); and reflexive monitoring (e.g. how do users appraise the intervention's value).

2.2. Study setting

Secondary care for people with chronic kidney disease in the UK is provided by over 70 main renal units; three of these acted as our study sites. We selected renal units as our study setting because they are more digitally mature than many other parts of the National Health Service. All renal units have an electronic patient record (EPR) system connected to a national infrastructure. Currently, this infrastructure—commissioned by the Renal Association—enables digital clinical data from renal units' EPRs to flow into a national repository to facilitate: audit by the UK Renal Registry; research; and patients' self-management via a patient portal (<https://www.patientview.org/>) (see Figure 1). As part of the OPT-ePRO intervention, we aimed to make this infrastructure bidirectional: with ePRO data—in our case: symptom burden and EQ-5D (i.e. quality of life) scores—flowing from the portal through the repository into renal units' EPRs.

2.3. Approach to intervention development

To develop the intervention, we used literature reviews [2,3] and frameworks [8,9] on implementing ePROs in general, complemented with context-specific qualitative data and stakeholder codesign (see below). We recruited participants from our study sites:

- Non-participant observations of workflows: we observed three 4-hour outpatient clinics (total: 12 hours) and eight in-centre haemodialysis sessions of nearly five hours each (total: 39 hours) to gain insight into existing routines and to identify opportunities for collecting and discussing ePROs. We used a data collection template to gather field notes on e.g. patient time and activities; staff presence and activities; and on how patient data was recorded, reviewed and discussed;
- Observations of clinic consultations: to understand if and how symptoms and quality of life were discussed, we audiorecorded and transcribed 11 routine clinic visits (177 minutes in total) in haemodialysis and outpatient settings;

- Participatory co-design workshops; we conducted three workshops with patients and caregivers (total n=25) and two with staff (n=13) to review designs and prototypes of intervention elements; identify potential implementation barriers; and to co-create solutions to overcome these through collaborative synthesis. Each workshop took approximately three hours.

Guided by our theoretical framework, we thematically analysed field notes, consultation transcripts and workshop materials. Using the constant comparative method [10] we synthesised results across sites, settings and data sources. We described intervention elements in line with relevant reporting guidance [11], while mapping them to NPT constructs. To deliver the IT-related intervention elements, we worked closely with study sites’ IT departments and the patient portal supplier. We provided them with technical and functional requirements (developed with input from end users and the Renal Association), as well as an end-to-end test plan and data.

The UK Health Research Authority’s Research Ethics Service, North West - Greater Manchester West Research Ethics Committee approved the study (ID 245870).

3. Results

The extended national infrastructure as outlined in Figure 1 encompassed all IT elements of our OPT-ePRO intervention. It enabled patients to enter ePRO data at home or in clinic via the existing patient portal, which they could access on any PC or mobile device. Many patients already had access to the portal, but only some were actively using it. Before each clinic consultation, staff would invite a patient to enter their ePRO data. Within 2-3 minutes, results were sent to the national data repository, and then pushed into the renal unit’s EPR system for clinical staff to review and discuss with the patient. An overview screen in the EPR displayed current and previous ePRO results in tabular format, with colour coding linked to symptom severity. When

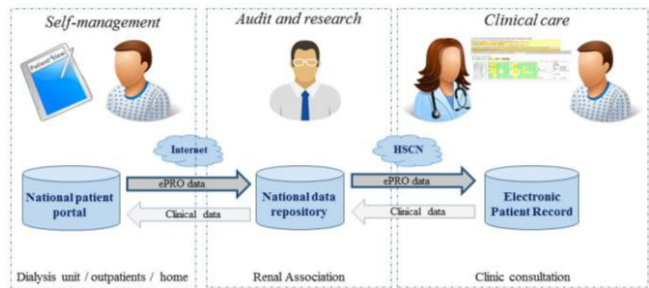


Figure 1 National infrastructure for capturing, storing, transferring and displaying ePRO data for people with kidney disease in the UK. HSCN refers to the NHS’ secure Health and Social Care Network

discussing results with patients, clinicians could generate pop-up screens to graphically display scores for individual ePRO items. There was a review functionality in the portal for patients to access their own ePRO results.

We identified relevant intervention components relating to values and roles, presented in Table 1 under the NPT constructs of coherence and cognitive participation. Components relating to service organisation and delivery are presented under collective action. We will explore reflexive monitoring –the process by which users adapt to and collectively assess the impact of the intervention—in future work evaluating the intervention in practice. Materials and procedures for each element left room for individual sites to organise the ePRO implementation in a way that fitted their local

context. All staff materials were combined into a handbook that contained information on how different aspects of the ePRO implementation would work. Patient materials were mostly delivered as flyers with concise messages, handed out in clinic or included in patient letters. Local champions were involved in delivering several parts of the intervention, and were usually a nurse manager and a consultant nephrologist.

Table 1. Overview and description of OPT-ePRO intervention elements, organised by NPT construct

Element name	Materials and procedures (what)	Delivery (how, by whom, where)
<i>COHERENCE & COGNITIVE PARTICIPATION</i>		
Implementation initiation and engagement	<ul style="list-style-type: none"> Initiating the implementation process and driving it until the local champion and other staff were ready to take this over Building relationships with and across all those involved, including patients, clinical staff, IT staff, management, volunteers, etc. 	F2F, email and phone conversations by study facilitator
Patient outreach	<ul style="list-style-type: none"> Explaining the implementation to patients; present potential benefits of ePROs; demonstrate ePRO module in patient portal Flyer announcing the implementation and listing potential benefits of ePROs 	<p>F2F conversations with individual patients in clinic by staff and volunteers</p> <p>Handed out in clinic or included in outpatient letters by staff</p>
Staff outreach	<ul style="list-style-type: none"> Planning implementation with local champion Explaining the implementation to staff; present potential benefits of ePROs Overview of the implementation process and of potential benefits of ePROs 	<p>Study facilitator</p> <p>F2F during team meetings by local champions and study facilitator</p> <p>Included in the handbook (on paper at nursing station and in electronic format)</p>
Assigning key staff roles	<ul style="list-style-type: none"> Identifying staff members or volunteers for key roles in collecting and discussing ePROs Checklist of implementation responsibilities, and the capabilities required to fulfill them 	<p>F2F conversations with individuals candidates by local champions</p> <p>Included in the handbook (on paper at nursing station and in electronic format)</p>
<i>COLLECTIVE ACTION</i>		
ePRO data entry & transfer	<ul style="list-style-type: none"> Providing patients access to the patient portal, i.e. registering new users and resetting passwords of non-active users ePRO data entry & transfer functionality for patients Patient materials, such as patient portal user guidance, password reminder cards, and a list of frequently asked questions regarding ePRO data collection and use 	<p>F2F conversations with individual patients in clinic by staff or volunteers; information included in outpatient letters by staff</p> <p>Screen accessible online by patients via the portal in clinic or from home; data transfer via the national infrastructure (Fig 1)</p> <p>Mostly on paper, handed out by staff in clinic; the list of frequently asked questions was accessible online by patients via the portal</p>
Review of ePRO results	<ul style="list-style-type: none"> ePRO review functionality for patients ePRO review functionality for staff Discussing ePRO results with patients 	<p>Screen accessible online by patients via the portal in clinic or from home</p> <p>Screens accessible by staff via the renal unit EPR system in clinic</p> <p>F2F consultations with individual patients by staff in clinic</p>
Staff training and support	<ul style="list-style-type: none"> Instructing key staff on how to fulfill their role in the ePRO implementation, including how to access and interpret ePRO results Training materials that outline the information provided during the F2F instruction sessions 	<p>F2F sessions with individual staff members in clinic by study facilitator and local champions</p> <p>Included in the handbook (on paper at nursing station and in electronic format)</p>

Abbreviations: EPR, electronic patient record; ePROs, electronic patient-reported outcomes; F2F, face-to-face

4. Discussion and conclusion

We developed a theory-informed and co-designed intervention to facilitate implementation of ePROs into UK renal services by supporting patients and staff with

embedding collection and use of this data into usual care pathways. Previous studies explored the feasibility of using tablets in renal settings to support ePRO collection locally [12,13], but our study is the first to deliver an infrastructure that is nationally scalable. In addition, we proposed a strategy to optimise use of ePROs in these settings, thereby addressing an acknowledged gap in the literature [14].

In the next phase of the OPT-ePRO study, we will evaluate the intervention in practice. When sites deploy the intervention, we will monitor for low ePRO response rates and conduct qualitative research to identify implementation barriers and to explore ways to address them. The qualitative research will also give us insight into how patients and staff understand and experience the intervention. This will inform iterative modifications of the intervention, mapped to the reflexive monitoring construct of our theoretical framework. Once all major barriers have been addressed, we expect the intervention to be suitable for deployment by other renal units, thereby enabling national implementation of ePROs in UK renal services and contributing to harnessing their potential benefits for patients and healthcare services.

Acknowledgments

The OPT-ePRO project is supported by the Health Foundation, an independent charity committed to bringing about better health and health care for people in the UK.

References

- [1] E. Basch, A.M. Deal, M.G. Kris, et al., Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. *J Clin Oncol.* (2016);34: 557–565.
- [2] A. Foster, L. Croot, J. Brazier, et al. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. *J Patient-Reported Outcomes.* *Journal of Patient-Reported Outcomes* (2018). 2: 1–16.
- [3] M.J. Reading, J.A. Merrill, Converging and diverging needs between patients and providers who are collecting and using patient-generated health data: an integrative review. *JAMIA.* (2018). 25: 759–771.
- [4] A. Nimmo, S. Bell, C. Brunton, et al. Collection and determinants of patient reported outcome measures in haemodialysis patients in Scotland. *QJM an Int J Med.* (2018). 111: 15–21.
- [5] A.A. Pagels, M. Stendahl, M. Evans, Patient-reported outcome measures as a new application in the Swedish Renal Registry: health-related quality of life through RAND-36. *Clin Kidney J.* (2019).
- [6] C.R. May, F. Mair, T. Finch, A. MacFarlane, C. Dowrick, S. Treweek, et al., Development of a theory of implementation and integration: Normalization Process Theory. *Implement Sci.* (2009). 4:29.
- [7] R. Mcevoy, L. Ballini, S. Maltoni, C.A.O. Donnell, F.S. Mair. A qualitative systematic review of studies using the normalization process theory to research implementation processes. *Implem Sci.* (2014). 9.
- [8] R. Gliklich, N. Dreyer, M. Leavy, Chapter 5: Use of patient-reported outcomes in registries. Registries for evaluating patient outcomes: a user's guide, 3rd ed. Rockville (MD): AHRQ; 2014.
- [9] I. Porter, D. Gonçalves-Bradley, I. Ricci-Cabello, et al. Framework and guidance for implementing patient-reported outcomes in clinical practice: Evidence, challenges and opportunities. *J Comp Eff Res.* (2016). 5: 507–519.
- [10] H.A. Boeije, A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Qual Quant.* (2002). 36: 391–409.
- [11] T.C. Hoffmann, P.P. Glasziou, I. Boutron, et al., Better Reporting of Interventions: Template for Intervention Description and Replication (TIDieR) Checklist & Guide. *Br Med J.* (2014). 348: g1687.
- [12] K. Schick-makaroff, A. Molzahn, Strategies to use tablet computers for collection of electronic patient-reported outcomes. *Health Qual Life Outcomes.* (2015). 13.
- [13] D. Wong, S. Cao, H. Ford, et al., Exploring the use of tablet computer-based electronic data capture system to assess patient reported measures among patients with chronic kidney disease: a pilot study. *BMC Nephrol.* (2017). 18: 1–10.
- [14] K. Schick-Makaroff, O. Thummapol, S. Thompson, et al., Strategies for incorporating patient-reported outcomes in the care of people with chronic kidney disease (PRO kidney): a protocol for a realist synthesis. *Syst Rev.* (2019). 8.

Development of a ePRO-Based Palliative Care Intervention for Cancer Patients: A Participatory Design Approach

Christina KARAMANIDOU^{a,1}, Christos MARAMIS^a, Kostas STAMATOPOULOS^a,
and Vassilis KOUTKIAS^a

^a*Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece*

Abstract. This paper describes a qualitative study conducted in the context of developing a novel ePRO (electronic Patient Reported Outcome) based palliative care intervention for cancer patients. The aim of the study was to elicit end-users' needs, judgements of the MyPal system and recommendations for improvement. A participatory design was chosen as the value of this approach has been well established in eHealth systems' design as well as the development of novel healthcare services. Focus groups with Chronic Lymphocytic Leukemia (CLL) patients were conducted at the Centre for Research and Technology (CERTH) in Greece using specially designed vignettes and discussion guides. Findings revealed that patients saw MyPal offering increased, direct contact with the healthcare team, freedom of physical and psychological symptom reporting as well as valid and reliable information. However, they had concerns about the appropriate use of data collected by MyPal, the efficiency of data analysis and data security adopted for sensitive personal information. The participatory design approach used has been very useful in encouraging the genuine involvement of participants, a factor which over time can empower and promote participants' long-term engagement.

Keywords. Palliative care, cancer, electronic patient reported outcome (ePRO), Participatory design, Focus Groups, eHealth

1. Introduction

There is a growing body of research demonstrating the value of participatory design in the development of novel healthcare services. In the context of palliative care, namely supportive care aiming to improve the quality of life of patients facing life-threatening illness through the assessment and treatment of psychosocial, physical and spiritual problems [1], participatory design is especially relevant. Considering that the demand for palliative care is increasing, discrepancies between care needs and care arrangements must be noted and rectified [2]. Actively involving patients in the design or development of eHealth tools [3], resources [4,5] or systems [6] for palliative care can provide a unique perspective on user acceptability, system usability as well as insight into the feasibility of the overall effort [7].

¹ Corresponding Author: E-mail: ckaramanidou@certh.gr.

MyPal (<https://mypal-project.eu/>) is a collaborative H2020 research project, funded by the European Commission, aiming to use eHealth technologies, in order to support cancer patients and healthcare professionals (HCPs). The main goal of MyPal is to develop and clinically assess new ePRO-based interventions for the palliation of cancer patients, in order to improve their Quality of Life (QoL) [8]. One group of cancer patients targeted by MyPal includes adults suffering from specific chronic hematologic malignancies. MyPal has committed to both adopting a patient-centered approach and adapting technology in order to cater for fundamentally different profiles of patients of different age groups as well as levels of digital and health literacy. Therefore, during the MyPal intervention design and protocol development, it is imperative to take into account not only the context i.e. current clinical practice for cancer patients, their interaction with healthcare professionals, provision of palliative care etc., but also users' personal needs as they identify and express them. To this end, use case scenarios were defined for the implicated stakeholders (patients and HCPs) to better illustrate the use of MyPal's technological tools appropriately responding to user needs in realistic situations. The aim of the present study has been to elicit the needs of end-users' representatives, their judgements with regards to different aspects (i.e. functionalities, characteristics) of the MyPal system, and finally their recommendations for system improvement.

2. Method

This was a cross-sectional study employing a qualitative methodology. Qualitative data were gathered via focus group discussions [9]. Ten focus groups (4-8 participants each) were conducted with: a) adult patients with a diagnosis of either Chronic Lymphocytic Leukemia (CLL) or Myelodysplastic Syndromes (MDS), and b) HCPs working in participating hospital units. Focus groups were conducted in partner sites in Greece (n=5; Italy (n=1); Germany (n=2), and the Czech Republic (n=2). This paper presents our overall methodology and the findings from the two focus groups conducted with CLL patients at CERTH, Greece.

With respect to materials used, two versions of vignettes were prepared: one for HCPs and one for adult patients. The following vignette, prepared for CLL patients, introduces the main character, Mr. Leonard Jones and his journey of managing CLL. The scenario starts with the enrollment of a treatment naïve patient, into the study. On Day 0 he completes the entry questionnaires, signs an informed consent and receives the MyPal smartphone application as well as a smart wristband. On Day 7, Mr. Jones receives a motivational message aiming to keep him engaged to the MyPal system. Two weeks later, he receives a notification to fill in the 1st symptom questionnaire via which he reports some fatigue, also is also becoming evident through a reduction in his physical activity. On Day 18, Mr. Jones searches for information in the personalized medical information search. He finds out that symptoms like fatigue combined with swollen lymph nodes can signify CLL progression. On Day 22 he develops rash and reports it via the spontaneous reporting functionality of MyPal app. He uploads a photograph and marks all his symptoms including swollen lymph nodes on a list. As all indicators (spontaneous reporting, symptom questionnaires and wristband data) demonstrate that CLL has progressed, MyPal issues an indication to the treating physician, who contacts Mr. Jones and asks him to undergo blood tests. After verification, on Day 24, Drug D is prescribed to Mr. Jones. He checks for drug

interactions and inserts timing preferences for reminder notifications on the MyPal app. Mr. Jones continues to receive motivational messages and self-report questionnaires to complete on physical and psycho-emotional symptoms. Mr. Jones keeps experiencing diarrhea, so on Day 42 he reports it via the MyPal app and his clinician is notified. As the diarrhea persists, he develops distress, evident in his facial expressions and tone of voice. Due to the distress picked up by the smartphone sensors in conjunction with the spontaneous reporting by Mr. Jones, the clinician is alerted and calls to reassure him. A sample screenshot from the vignette under discussion is presented in Figure 1.

The discussion guides reflected the episodes of the vignettes from enrollment up until the end of the storyline. An example follows: *‘Mr. Jones starts feeling fatigue and searches for info, since he thinks this might indicate disease progression. The information that he finds via MyPal is personalized, i.e. information relevant to him as a CLL patient. How do you judge this feature? How would you feel given the opportunity to access information tailored for you as a patient with CLL? How can it be improved?’*.

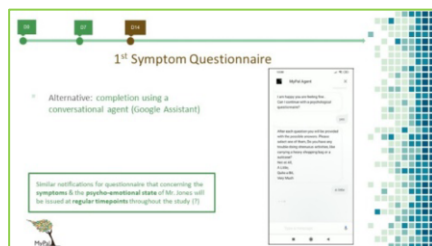


Figure 1. Sample screenshot from the patient vignette.

Concerning data analysis, the framework method was employed [10] for performing thematic analysis. It is an established approach for managing and analyzing qualitative data in applied health research [11]. Ensuring uniformity of the focus group analysis in the different participating sites was a primary concern. The main reasons for choosing this method of analysis included: a) the multi-disciplinary research teams involved, b) the sample with clinical and patient representation, c) the complexity of the collected data pertaining to 4 different countries and languages, and d) the nature of the data consisting of expressed positive and negative judgements as well as concrete recommendations for MyPal system development. The collection of data in diverse cultural contexts and from multiple perspectives contributed to the richness and thickness of data. Codes were developed at a national level and these were grouped into categories at a European level. Researchers are satisfied that data saturation has been achieved per emerging category for both clinical and patient groups.

User acceptance is not only a methodologically major challenge, but also one of significant ethical importance as it aims to enhance patients' autonomy and is directed towards their beneficence [12]. MyPal is indeed tailoring the technological aspects of the intervention to the exact patient needs, rather than requiring patients to adapt themselves to the employed technology. During recruitment participants were informed about the study, gave their consent for participation, recording of the discussion and use of content for the needs of the MyPal program. The information collected did not include any personal data but rather information on the functionality of the system as well as recommendations for better use. It must be noted that in the clinical studies to

follow later in the course of this project, consent forms will be submitted with the actual protocol to the appropriate Research Ethics Committees for approval.

3. Results

Nine patients with CLL, aged 39-73, participated in the two focus groups held at CERTH, Greece. Five were males and four were females. Three participants were receiving treatment at the time of the focus group, four had received treatment in the past, while two were treatment-naïve. CLL patients expressed what they considered to be the most important advantages that the MyPal system has to offer (Table 1), but also voiced important concerns (Table 2). Participants have given recommendations which have led to amendments in the MyPal system design and MyPal intervention delivery. Indicative examples can be found in Table 3.

Table 1. Patients’ positive judgements on aspects of the MyPal system.

Themes	Explanation
Valued contact with the healthcare team	MyPal promotes direct contact with the doctor, instantly updating the medical file and, thus, expediting medical decision-making and initiating supportive treatment following the presentation of an issue the patient is concerned about.
Freedom of reporting	MyPal allows the communication of physical and psychological symptoms attributed to either the disease or its treatment through ePROS or other information difficult for patients to assess. Patients positively perceived MyPal’s capability to record on a scheduled or spontaneous basis, in a structured or unstructured way.
Validity and reliability of information	The MyPal database will relieve them from stress and put a stop to misinformation. It will be safer and much more specialized than most generic databases available on the web since many physicians will be contributing throughout the study.

Table 2. Patients’ concerns regarding aspects of the MyPal system.

Themes	Explanation
Reliability of collected data	The data generated might not provide essential or reliable information (e.g. the data on the quality of sleep might be affected by the use of sleeping aids).
Interpretation of collected data	The data gathered might not be interpreted correctly (e.g. smartphone sensors might not be able to distinguish real distress from happiness and exertion). The need to combine different sources of information, so that data is interpreted in context was stressed.
Privacy and digital safety of collected data	Patients desired transparency over issues such as who will be viewing their personal files, who these data belong to, where data will be stored, etc. Finally, concerns were voiced about safeguarding privacy, e.g. whether the system will be recording information without patient consent and/or awareness (e.g.. smartphone sensors reading facial expressions and tone of voice).

Table 3. Impact of participants’ recommendations on MyPal system design and intervention delivery.

MyPal proposal	Participants’ feedback	Amendment Decisions
MyPal will notify participants that a PRO assessment is due and accept submitted PRO reporting	Patients stressed the need for feedback after submission	The MyPal system will issue a notification to inform the participant that the information has been successfully submitted and that it will be seen by a HCP as soon as possible.
The MyPal system was designed to record spontaneous symptom reporting and scheduled symptom reporting	Patients expressed concerns that the physician will not see the information in real time or that he/she will not be available or not respond	Informed consent will state that MyPal is not an emergency service. During symptom reporting, patients will be informed that in case of emergency they should contact the hospital directly. Participating HCPs will check the system for new information every 72 hours.

4. Lessons learnt

The purpose of this study was to present the vision of the MyPal system to all implicated stakeholders. Findings from the focus groups conducted with CLL patients offered insights into patients' requirements from an eHealth system like MyPal. It was evident that, in order to support patients and their caregivers in getting the most of MyPal, it is necessary to alleviate their concerns about the appropriate use of data and the efficiency of data analysis, while also providing reassurance with high standards of data security adopted for sensitive personal information. End users' valuable contributions are already informing and shaping the MyPal design and the next cycle of iteration has already been planned for preliminary field testing. The employed participatory design approach has been very useful in that it has encouraged genuine involvement of participants, a factor which over time can empower and promote participants' long-term engagement.

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825872.

References

- [1] C. Sepúlveda et al., Palliative care: the World Health Organization's global perspective, *J Pain Symptom Manage.*, **24** (2002), 91–96.
- [2] B. Russell et al., Responding to urgency of need: Initial qualitative stage in the development of a triage tool for use in palliative care services, *Palliat Med.* **32** (2018), 1246–1254.
- [3] M. J. Gonzales, & L. D. Riek, Co-designing patient-centered health communication tools for cancer care. In Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare (2013), 208–215.
- [4] D. Scott, et al., Development of an eHealth information resource for family carers supporting a person receiving palliative care on the island of Ireland. *BMC palliative care*, **18**(1) (2019), 74.
- [5] A. F Street et al., Developing a web-based information resource for palliative care: an action-research inspired approach. *BMC medical informatics and decision making*, **7**(1), (2007), 26.
- [6] M. J. Allsop et al., Multidisciplinary Software Design for the Routine Monitoring and Assessment of Pain in Palliative Care Services: The Development of PainCheck. *JCO Clinical Cancer Informatics*, **3** (2019), 1–17.
- [7] M.E. Nieroda et al., Online Decision Support Tool for Personalized Cancer Symptom Checking in the Community (REACT): Acceptability, Feasibility, and Usability Study, *JMIR Cancer* **4** (2018), e10073.
- [8] C. Maramis et al. Using Electronic Patient Reported Outcomes to Foster Palliative Cancer Care: the MyPal Approach, in: Proc. of the 19th IEEE Int. Conf. on Bioinformatics and Bioengineering, IEEE, Athens, Greece 2019 – to appear.
- [9] S.A. Fricker et al., Requirements Engineering for Digital Health, Springer, Cham, 2015.
- [10] N.K. Gale et al., Using the framework method for the analysis of qualitative data in multi-disciplinary health research, *BMC Med. Res. Methodol.* **13** (2013), 117.
- [11] J. Ritchie et al., Qualitative Research Practice: A Guide for Social Science Students and Researchers, SAGE, 2013.
- [12] B. Monroe Oliviere D. Patient Participation in Palliative Care: A Voice for the Voiceless. Oxford University Press, 2003, ISBN 0-19-851581-2.

Development of a Mobile Digital Manikin to Measure Pain Location and Intensity

Sabine N VAN DER VEER^{a,1}, Anna L BEUKENHORST^b, S Mustafa ALI^b, Ben JAMES^c, Pedro SILVA^c, John McBETH^b, William G DIXON^b

^a *Centre for Health Informatics, University of Manchester, Manchester, UK*

^b *Centre for Epidemiology Versus Arthritis, University of Manchester, Manchester, UK*

^c *uMotif, London, UK*

Abstract. Painful conditions are prevalent and substantially contribute to disability worldwide. Digital manikins are body-shaped drawings to facilitate self-reporting of pain. Some of them have been validated, but without allowing for recording of location-specific pain intensity and for use on a smartphone. This paper describes the initial development of a digital pain manikin to support self-reporting of pain location and location-specific intensity using people's own mobile device. Subsequently, we conducted reliability and usability tests with eight researchers and seven patient representatives. Test-retest reliability depended on the manikin's level of detail, but was generally high with most intraclass correlation coefficients ≥ 0.70 and all similarity coefficients ≥ 0.50 . Participants found the manikin easy to use, but suggested clearer orientation (front/back, certain body locations) and would value additional feedback and diary functions. We will address these issues in the next version of the manikin before conducting a validation study.

Keywords. Pain measurement; Patient-generated health data; mHealth

1. Introduction

Pain is one of the leading causes of disability worldwide [1]. It is a symptom of all musculoskeletal diseases, and prominent in many other conditions, such as cancer. Pain substantially affects people's quality of life and imposes a significant burden on society due to treatment costs and absence from work [2]. To optimally manage pain and understand treatment response, we need insight into how it changes over time. This requires validated instruments to capture key aspects of pain, which often vary per location, such as pain intensity.

Digital manikins are human-shaped drawings where people can shade areas to self-report the location of their pain on a computer. In contrast to their paper counterparts, digital manikins support automated (rather than manual) calculations of e.g. the extent of the affected area. Previous studies presented validated, digital pain manikins [3–5], but none of them facilitated the capture of location-specific pain intensity (i.e. different intensities for different locations within one pain report). They also required a dedicated device with a pre-defined screen size, limiting opportunities for people to use their personal devices, which in turn hampers deploying these manikins for large

¹ Corresponding Author; E-mail: sabine.vanderveer@manchester.ac.uk

mHealth studies. Manikins incorporated in pain self-management apps that are currently available in app stores (e.g. <https://www.catchmypain.com/>) are unsuitable as research tools because they have not been validated as measurement instruments. We, therefore, developed a digital pain manikin to enable people to self-report the location and location-specific intensity of their pain using any (mobile) device. This paper reports its initial development and usability testing.

2. Methods

2.1. Initial development

We developed a prototype, for Android IOS in the first instance, together with an industry partner (<https://www.umotif.com/>). The prototype design was guided by the manikin's intended purpose, published digital pain manikin studies [3–5], and input from clinicians and researchers with pain measurement expertise. The manikin was two dimensional, black, and gender-neutral with anchors for chin/back of the head and groin/buttocks; users could select a zoomed-in view on the lower arm and hand. There was a slider to indicate pain intensity before shading an area; score and colour ranged from 1 (grey; no pain) to 10 (red; severe pain). Users could erase incorrectly shared areas by selecting an intensity of 1. There was a visible grid with squares overlaying the manikin; 'active' squares were those that had been shaded with intensity >1. A mechanism for exporting a table with the coordinates and intensity scores of all active squares facilitated a pain score calculation that was not influenced by screen size.

2.2. Selecting the optimal grid

As we expected the grid's granularity to affect both user experience and the accuracy of pain drawings, we evaluated the usability and reliability of three grids with different granularities: coarse (40x80 squares); medium (80x160); and fine (100x200); we tested full body and zoomed-in arm manikins separately. We aimed to identify the grid that best balanced ease of use with drawing accuracy.

We recruited eight researchers from our professional network; five had experience with measuring pain. A sample size of eight was sufficient to detect an intraclass correlation coefficient (ICC) of 0.70 or more [6], found by previous studies [3–5]. After receiving verbal instructions and completing one practice drawing, participants completed a first test set of six drawings (i.e. one for each grid-manikin combination) on a Samsung S6 smartphone by copying an example manikin from paper (Figure 1); we randomised the order in which grids were tested. We then interviewed participants for 30 minutes to explore their views on the ease of use of each grid before asking them to complete a second set of the same six drawings (i.e. retest).

For each grid-manikin combination we evaluated the test-retest reliability of pain extent by calculating (i) the size (i.e. mean of difference) and range (i.e. limit of agreement as $\pm 1.96 \times \text{standard deviation}$) of within-participant differences between test and retest [7] for the proportion of active squares within the manikin perimeter, and (ii) the ICC [8]. For the test-retest reliability of pain location we calculated the Jaccard similarity coefficient (per [3]), which is the number of squares that were active in both test and retest divided by the number of squares that were active in either test or retest; values range from 0 to 1 (no to complete overlap in pain location). We also calculated mean completion time during the retest and synthesised interview notes per grid.

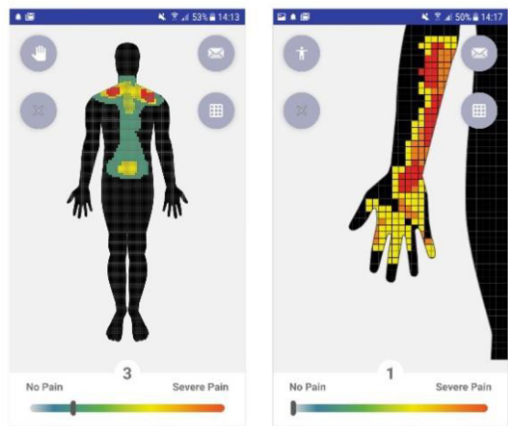


Figure 1. Screenshots of the medium grid body manikin (left panel) and arm manikin (right panel), as completed on a Samsung S6 smartphone by one of the participants

2.3. Exploring usability

We explored the general usability of the digital pain manikin, including the manikin’s ease of use; its look and feel; and suggestions for additional functionalities to improve its usefulness for pain management more widely. We collected qualitative data through the researcher interviews (see 2.2), complemented with a 3-hour workshop with seven members of a patient and public involvement (PPI) panel with personal experience of chronic pain. The workshop included individual 30-minute sessions where participants were interviewed and performed a series of tasks with the manikin using the medium grid. They then shared their experiences in a group discussion. We imported written interview and workshop notes into NVivo and analysed them thematically.

3. Results

3.1. Grid selection

Table 1 shows that for pain extent, the coarse grid had the lowest test-retest reliability for both manikins with the highest mean of difference and the lowest ICC. The medium grid performed best for the body manikin, with the lowest mean of difference and the highest ICC. The fine grid showed the best results for the arm manikin. For pain location, the test-retest overlap was comparable between grids for the body manikin. For the arm manikin, the coarse grid showed most overlap, but also the highest proportion of active squares in test and retest. Within grids, completion time varied widely between participants but generally increased with granularity.

Compared to the fine and medium grids, participants found the coarse grid easier to use but also felt it produced the least accurate pain drawings. Others reported a higher risk of unintentionally activating squares, and having to dab (rather than draw) to shade areas. In contrast, they considered the fine grid the most time-consuming, but also the most accurate method that felt most like drawing. Some participants questioned whether real pain reports would require such a fine level of detail. For the

medium grid, most participants reported similar experiences as for the fine grid, although some explicitly stated that the medium grid better balanced ease of use with drawing accuracy.

Table 1. Test-retest reliability of pain extent and location, and completion time per grid-manikin type combination

Manikin type →	Coarse grid (40x80)		Medium grid (80x160)		Fine grid (100x200)	
	Body	Arm	Body	Arm	Body	Arm
<i>Extent</i>						
Number of available squares ^{a)}	1,866	133	6,922	483	10,519	734
Mean (SD) proportion of active squares ^{b)}						
Test	0.17 (0.09)	0.39 (0.09)	0.16 (0.07)	0.34 (0.06)	0.14 (0.08)	0.34 (0.10)
Retest	0.16 (0.06)	0.44 (0.07)	0.13 (0.06)	0.36 (0.05)	0.14 (0.05)	0.37 (0.07)
Mean (1.96×SD) of difference ^{c)}	4.39 (5.13)	5.83 (12.62)	2.31 (3.90)	4.48 (6.92)	2.60 (6.72)	3.90 (7.98)
ICC (95% CI)	0.77 (0.27 – 0.95)	0.48 (-0.23 – 0.87)	0.90 (0.63 – 0.98)	0.51 (-0.20 – 0.87)	0.81 (0.36 – 0.96)	0.80 (0.33 – 0.96)
<i>Location</i>						
Mean (SD) Jaccard coefficient	0.50 (0.15)	0.71 (0.12)	0.51 (0.16)	0.59 (0.12)	0.54 (0.15)	0.59 (0.13)
<i>Completion time (retest)</i>						
Mean (SD) number of seconds	64.0 (57.4)	47.5 (21.7)	72.0 (52.2)	68.2 (40.7)	82.1 (33.7)	70.1 (42.2)

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; SD, standard deviation
^{a)} Total number of squares in the grid that fall within the manikin perimeter
^{b)} Proportion of available squares shaded with pain intensity >1
^{c)} Difference and limit of agreement between test—retest [7] in the proportion of active squares

3.2. Usability of the manikin

All participants found completing a manikin report easy and expected basic user instructions and some practice to be sufficient for most people to use it unsupervised. Some participants made sweeping movements to activate squares, while others dabbed, especially PPI members. Dabbing was more time consuming, which may explain why some PPI members would prefer a coarser grid. Many participants ‘pinched’ in an attempt to zoom in on certain body locations, but were equally happy to use the predefined zoomed-in views. Some participants struggled to activate squares as intended, especially on the border of the manikin perimeter. The method for deactivating squares was considered acceptable, although some suggested an eraser or ‘undo’ function.

Most participants saw the manikin as male, and PPI members additionally characterised it as “elongated” and “a frog man”. Additionally, many had trouble distinguishing front from back and therefore suggested adding labels (front/back, left/right), as well as more and clearer anchors (e.g. chest, lower back). They generally liked the grid being visible, and most considered the colour scheme for pain intensity appropriate. However, some found green colours for scores of 4-5 counterintuitive and many would prefer an intensity score of ‘0’ (rather than ‘1’) to reflect ‘no pain’. No one expressed concerns with the manikin being black.

Participants suggested several additional functionalities for improving the manikin’s usefulness for pain management, including: option to record additional pain aspects (e.g. location-specific pain type (throbbing, stabbing, etc.)), pain interference with daily life); feedback of pain reports over time (e.g. heat maps); diary to record free text information alongside a manikin report (e.g. description of the pain experience); and an option to share manikin reports with your doctor to inform (tele)consultations.

4. Discussion and conclusion

We developed and tested a digital pain manikin that facilitates measuring self-reported pain location and location-specific intensity on any device. Study participants found the manikin easy to use, and it generally showed moderate to very high reliability.

A limitation of our study is that we asked participants to copy a detailed manikin example, which might be more complex than reporting a personal pain experience. They also had to do this 2×6 times instead of the usual 2×1. The increased task complexity and repetitiveness may explain why our ICCs for pain extent did not exceed the 0.90 reported by other studies [3–5]. The reliability of pain location seemed less affected by this, with Jaccard similarity coefficients in some other studies not exceeding 0.50 [3,5].

Overall, the fine grid came out as most reliable across manikins, but the usability tests suggested that it might be too difficult and time consuming to use. The next version of the manikin will, therefore, use the medium grid, as well as addressing most of the issues and suggestions this study identified, such as: providing written user instructions; improving the look and feel of the manikin; facilitating the capture of additional pain aspects, and including diary and feedback functions. Through continued collaborations with patients, clinicians, researchers and industry, we will use the new manikin version to collect digital pain reports in a range of painful conditions to (i) reassess the manikin's reliability for capturing people's personal pain experience, and (ii) develop and validate algorithms for calculating quantitative manikin scores that combine pain location, extent and intensity. Ultimately, we expect this contribute to our digital pain manikin becoming a new standard for digital pain reports in large mHealth studies, while also being suitable for supporting clinical care and self-management.

Acknowledgments This project was supported by the Centre for Epidemiology Versus Arthritis (ref 20380).

References

- [1] S.L. James, D. Abate, K.H. Abate, et al., Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017, *Lancet*. (2018). doi:10.1016/S0140-6736(18)32279-7.
- [2] Institute of Medicine, Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research, The National Academies Press, Washington, DC, 2011. doi:10.17226/13172.
- [3] M. Barbero, F. Moresi, D. Leoni, R. Gatti, M. Egloff, and D. Falla, Test-retest reliability of pain extent and pain location using a novel method for pain drawing analysis, *Eur. J. Pain (United Kingdom)*. (2015). doi:10.1002/ejp.636.
- [4] D. Leoni, D. Falla, C. Heitz, G. Capra, R. Clijsen, M. Egloff, C. Cescon, J.P. Baeyens, and M. Barbero, Test-retest Reliability in Reporting the Pain Induced by a Pain Provocation Test: Further Validation of a Novel Approach for Pain Drawing Acquisition and Analysis, *Pain Pract*. (2017). doi:10.1111/papr.12429.
- [5] M.A. Bujang, and N. Baharum, A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review, *Arch. Orofac. Sci.* (2017).
- [6] J.M. Bland, and D.G. Altman, Measuring agreement in method comparison studies, *Stat. Methods Med. Res.* (1999). doi:10.1177/096228029900800204.
- [7] T.A. Neubert, M. Dusch, M. Karst, and F. Beissner, Designing a tablet-based software app for mapping bodily symptoms: Usability evaluation and reproducibility analysis, *JMIR MHealth UHealth*. (2018). doi:10.2196/mhealth.8409.
- [8] S. Chinn, The assessment of methods of measurement, *Stat. Med.* **9** (1990) 351–362.

Digital Support for Self-Management in Children with Diabetes: Understanding Their Needs and Developing a Design Concept

Nicole Safavi ANDERSEN^a, Louise Hedrich HAUGAARD^a, Sissel Bøgh PEDERSEN^a, Mikkel Strange PEDERSEN^a & Ann BYGHOLM^{a,b} ¹

^a*Department of Communication and Psychology Aalborg University Denmark*

^b*Danish Centre of Health Informatics, Aalborg University Denmark*

Abstract. In this paper we present a study on the needs and requirements in a digital tool to support self-management in children and adolescents (0-18 years) with type 1 diabetes. The study was formally initiated by the Steno Diabetes Center North Jutland (SDCN), which wanted to develop a virtual universe that could support children and young people in coping with their disease. Through this study we interviewed the project manager from SDCN and the head of the family group in North Jutland. Furthermore, we observed a workshop involving health care professionals and learning specialists who were working with ideas for the virtual universe. We conducted two focus group interviews, one for children with type 1 diabetes and one for parents with diabetic children. The analysis of the data revealed very different ideas about what digital support is important for the self-management of this target group. Health care professionals focused on developing various learning materials using virtual reality or augmented reality, while the parents desired security and control and requested materials that could inform and educate people in their locale, such as school teachers, parents of their children's friends, etc. The children emphasized the importance of not being singled out and requested simple tools that could be used from their smartphones. They especially asked for tools that could help them manage their diabetes by themselves in a fast and effective way, e.g., making it easier to count carbohydrates and calculate their insulin intake. Based on the analysis, we developed suggestions for a design. The main lesson learned from this study was the importance of focusing on the *actual* users and their everyday lives in developing new tools, and not on technological possibilities.

Keywords. diabetes; app design; user research;

1. Introduction

In 2017, approximately 425 million adults (20-79 years) and more than 1,106,500 children lived with the chronic disease diabetes [1]. The most well-known types of diabetes are type 1 and type 2. Type 1 diabetes occurs when people cannot produce insulin themselves and type 2 diabetes occurs when people cannot use insulin to turn glucose into energy [1]. The study presented in this paper focuses on children diagnosed with type 1 diabetes [2].

¹ Ann Bygholm, ann@hum.aau.dk,

Diabetics have an increased risk of serious health issues and complications, including cardiovascular disease, blindness, kidney failure, and lower limb amputation [3]. It is therefore important that children and young adults learn to self-manage and take responsibility for their disease even at a young age. To help children and young adults manage their diabetes, the Steno Diabetes Center North Jutland (SDCN) wished to develop a digital concept that supports diabetic children and their families in their everyday life. When we began our cooperation with SDCN, the vision for the project was still in a very early state, but SDCN knew they wanted “something digital” and were interested in the idea of a virtual universe for the children. It was important that the digital device would support self-management and empowerment, that is, to maximize the self-care knowledge, skills, self-awareness, and sense of personal autonomy of patients and to enable them to take charge of their own care [4].

To understand the context of diabetes-related technology, we explored existing ICT apps and tools for diabetes on the market. We found several diabetes-related apps, but almost all the apps were made for monitoring diabetes. We found only one app, provided by the Danish Diabetes Association, that was developed to help children master and manage diabetes. Monitoring and mastering/managing are different types of help: monitoring focuses on keeping track of your numbers (insulin, blood glucose levels, etc.) while self-mastering also focuses on having the knowledge and competences to handle and manage the disease. The collaboration with SDCN led to a study on the needs and requirements for digital support to aid self-mastery for diabetic children, aged 10-18 years, in order to develop suggestions for a design.

2. Methods

For this study, we applied a user-centered approach [5] to understand the aims and needs of the stakeholders: the children, the parents, and the health care professionals.

We conducted an unstructured interview with the project manager from SDCN with the aim of understanding the thoughts and wishes regarding a virtual platform for children and teenagers and why this project was initiated. A few days after this interview, we participated in a SDCN workshop to inspire innovation in digital health for children and teenagers with type 1 diabetes. Doctors and nurses from Aalborg University Hospital worked with two digital learning specialists from a company called CadPeople and the project manager from SDCN. While observing this workshop, we conducted an unstructured non-participatory observation and took fieldnotes. After collecting data from health personnel and experts, we wanted to obtain the users' perspective. Therefore, we had a conversation with the chairman of the Danish Diabetes Association of Northern Jutland. Through our contact with the chairman, we arranged two separate focus-groups, one with 8 children, aged 9-13 years, and another with 7 parents. We prepared an interview guide for the children with questions that allowed us to focus on their experiences and feelings about being diagnosed with diabetes and dealing with the disease. We asked the parents questions about parenting a child with diabetes. As expected, the parents were able to reflect upon the questions, whereas the children needed help. Therefore, we chose to include the generative techniques inspired by Sanders and Stappers [6]. We used pictures in the form of different social activities to provoke reflection and help the participants to recall memories, make interpretations, explain feelings, and imagine future experiences.

Data from the interviews, observations, and focus groups were transcribed in order to for us code it and draw out the relevant knowledge from the study. The coding process was approached with an open mind; as a framework we followed Braun and Clark's six-step thematic coding [7]. This led to the identification of important themes that are presented in the findings below.

3. Findings

The identified themes represent the different issues and challenges that the children and their parents encounter on a daily basis, including "Practical challenges with diabetes," "Challenges when educating the network," "Lack of knowledge about the illness," "The balance between independence and childhood," and "Overprotective parents". We condensed the themes further and ended up with three main concerns that should be addressed in a digital tool that could support self-management. The first one is easy access to carbohydrate calculation, the next is easy access to relevant guides and information, and the third one is what to do in an emergency.

Besides these findings, it was also clear from our interview with the children that they wanted a solution that they could carry with them all the time. The children already have access to an app that counts carbohydrates, provided by the Danish Diabetes Association. But this app does not provide its users with sufficient options or a large enough database to support self-management. Inspired by this app, our aim was to come up with a digital solution that improves all aspects and functionalities, based on the identified themes and concerns from our analysis.

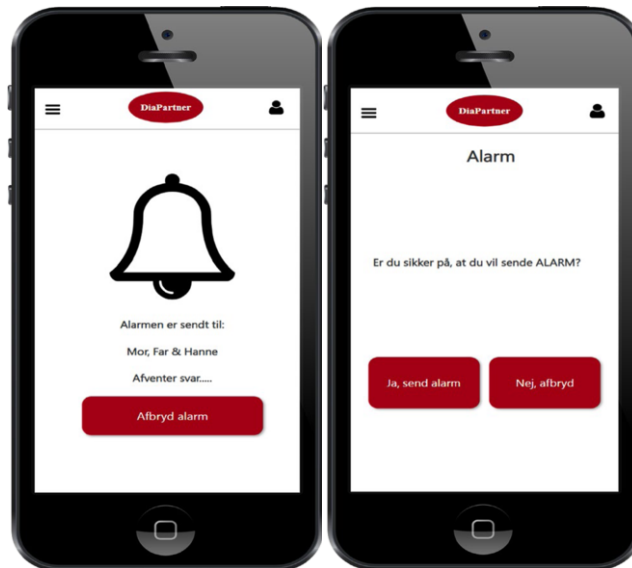
The current app is not easy or fast to use in counting carbohydrates. For example, the application should provide the users with a larger database of food and beverages that must be easy to navigate and the users must be able to select specific kinds of food and beverage as "favorites" (see Picture 1).



Picture 1. Show product and save as a "favorite"

Furthermore, the application had to inform and provide advice about type 1 diabetes, not only to enlighten its users (the children with diabetes), but also to provide the network with an easy and accessible way to learn about the disease. Thus, it was a main concern for the parents that e.g. school teachers, sports association leaders and others did not have the necessary knowledge about diabetes to take care of their child. Therefore, the child could not participate in social events like summer camps and sleepover parties. An important part of this issue is also a specific guide on how to handle a hypoglycemic incident

Another result of the data analysis showed that there is a need for a function that could help the children and their parents to feel more confident that they will get help quickly if needed, especially when the children are away from home. Therefore, we decided to incorporate an alarm. To do so, we had to include and activate the phone's GPS, which caused several ethical questions whether the purpose of the application is to support the children's self-management—not for the parents to be able to track them. Instead of activating the GPS all the time, we developed the ALARM function so the users (children with diabetes) could include three different emergency contacts. If the alarm is triggered, these contacts receive a message with the child's location (see Picture 2).



Picture 2. ALARM

In summary, the design proposals point to a relatively simple solution that particularly addresses the demands of children for easy access to carbohydrate calculations and the parents' requirements for security and access to relevant information for networks and friends.

4. Discussion

The analysis of the data from the health professionals, parents, and children revealed very different foci about what is important in digital support for children and adolescents

with type 1 diabetes. The workshop with health care professionals resulted in three design suggestions: a digital diabetes diary, an interactive understanding of diabetes, and the implementation of virtual reality to build bridges between parents and children. The parents' concerns were connected to the security of their children; for example, "how much responsibility should be given to a 7-year-old?" and "are the teachers aware of how to help and support my child?" Also, the parents assumed that the children would like to participate in an online community. The children, however, were not interested in digital diaries, augmented reality, or online communities. They wanted an optimization of a carbohydrate app that could help them manage their diabetes, provide security when on their own, and offer information when needed. As stated above, our study shows the importance of putting the end-users and their daily life situation at the center when designing new applications.

References

- [1] International Diabetes Federation. (2017). About diabetes. Retrieved May 7, 2019, from <https://www.idf.org/aboutdiabetes/what-is-diabetes.html>.
- [2] Andersen, N. S., Pedersen, S. G. B., Haugaard, L. H., & Pedersen, M. S. (2019). *DiaPartner—Your diabetic toolbox* (p. 72) [A project about patient empowerment and supporting diabetic children and teenagers in their everyday life]. Aalborg University.
- [3] International Diabetes Federation. (2018). Diabetes Complications. Retrieved May 7, 2019, from <https://www.idf.org/aboutdiabetes/what-is-diabetes/complications.html>.
- [4] Anderson, R. M., Funnell, M. M., Barr, P. A., Dedrick, R. F., & Davis, W. K. (1991). Learning to empower patients. Results of professional education program for diabetes educators. *Diabetes Care*, 14(7), 584-590.
- [5] Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, 5-18. <https://doi.org/10.1080/15710880701875068>.
- [6] Sanders, L., & Stappers, P. J. (2012). Chapter 3: How it works—Generative tools and techniques. In *Convivial design toolbox: Generative research for the front end of design* (pp. 65–96). Amsterdam: BIS.
- [7] Braun, V. & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3 (2): 77-101. <https://doi.org/10.1191/1478088706qp063oa>.

Does Health-IT Improve People Centered Care?

Pernille BERTELSEN¹, Lone Stub PETERSEN¹ and Christian NØHR²

¹Aalborg University, Department of Planning, Danish Centre for Health Informatics,

²Maersk Mc Kinney Moller Institute, University of Southern Denmark, Denmark

Abstract. Patient involvement can be supported in many ways, some of these through health IT (HIT). Health care practice is changing with digitalization and technologies that patients can use to generate data outside the formal health facilities are growing. Inequality affect citizens ability to participate. There is a need to be mindful of the perceptions and the willingness of citizens and the health care professionals to meet this change. Selected data from two Danish surveys, highlight the citizens and health care professional's perspective on health IT. From the analysis emerge central considerations on what to be aware of in the future development of HIT.

Keywords. eHealth, Survey, Citizens, Clinicians, Equality.

1. Introduction

1.1. Strategy and practice

In line with the WHO global strategy on people-centered and integrated health service [1] the latest Danish Digital Health Strategy 2018-2022 emphasize that: "Patients should be more involved in their own treatment, so that the health system interacts with the patients on his/her own terms" [2] (p.21). Throughout the strategy we can read an increased focus on the involvement of the patients and an interest to increase the support to them and their relatives "... in taking greater ownership of their own illness in their daily lives ..." [2] (p.18).

There is a general international interest in the use of what is termed patient generated health data (PGHD), in the healthcare sector [3]. The availability of technological solutions makes it possible for citizens and patients to collect data themselves. This is one of the ways of framing patient involvement through HIT. However, equality and inequality aspects of HIT is not predominant.

The strategy recognizes that in the formal health care system there are relatively few experiences using data collected by the patient's own equipment but also "that the areas where apps, medical devices and medication converge are currently shifting" [2] (p.24). Hence there is a need to develop a shared framework that allow patients to start using devices and apps and share their data with the health care system. A central notion is to combine the potential of HIT and PGHD to improve service and reduce costs. However, in Denmark and in most other countries little is known about what type of technologies citizens use, and which social groups use the technologies as well as how clinicians and

citizens recognize and use digital health care [4]. Here we explore the different perspectives on the use of patient generated health data from selected questions in two different surveys.

Present-day, we see small devices sensing a range of vital measures among average consumers allowing them to track their own health data. It is no longer exclusive toys for more well-off first movers [3]. Although the potential in PGHD and health apps for mobile devices is immense a German study shows that a substantial proportion of the population is not engaged [5]. They conclude that although there are many health app users a substantial proportion of the population are not engaged. They suggest that socio-economic-related and literacy-related disparities as well as motivation to change should be considered by app developers and researchers. A Study from US show that Collage education has a significant positive impact on eHealth behaviors [9]. Nonetheless, as stated in the Danish digital health strategy, it still remains to work out how best to integrate these data into contemporary clinical work, and as emphasized by [5], to make sure those in most needs also are those being encouraged to collect own health data and thus benefitting from the use of the health data.

To direct the future development of citizens centered health apps it is necessary to have more knowledge on who the actual users of health apps are, and knowledge on the health providers attitudes towards citizens using these tools and services.

2. Methods

Selected questions from two different studies were used to analyze the citizens' experiences with accessing and using health information systems, and health providers attitudes to opportunities for patients' use of information systems and the impact of these systems on patients' engagement in personal health.

2.1. Citizen survey

The 2017 citizen survey reported from in this paper is the third in a bi-annual series of National surveys on Danish citizens' expectations and perspectives on eHealth. A questionnaire was developed by the authors and a Danish market research agency (Megafon), was commissioned to administer the surveys, on a population sample of $n=1059$. The survey is combinational using both email and telephone. The selected respondents are part of a citizens' panel reflecting the Danish adult population with respect to age, education and geographic distribution. The questionnaire consists of 10 demographic background questions (Age, gender, education, state of health, etc.) and 39 questions about their relations and attitudes towards their use of e-health technologies (hardware, service used, etc.). Data from three questions are reported in this paper.

2.2. Clinician survey

The 2017 study of clinicians use of health information systems has been conducted for 9 consecutive years. Questionnaires was distributed to medical doctors, nurses and medical secretaries with the help from their respective professional associations. Random samples (in average 10.000) from their member files received a link to a web-based questionnaire. An average response rate of 23% was achieved over the 9 years of study. The questionnaire consists of 7 background questions (Age, gender, education, seniority,

medical specialty etc.) and 48 questions about their use of specific systems, their user experience (Ux) and attitude towards e-health in general. Data from one question is reported in this paper.

3. Results

3.1. What is the educational background of the citizens using e-health?

There is a clear majority of citizens with a master level university education that are using e-health services to monitor their health or seek health advice (Figure 1). 48% were familiar with one or more of these services compared to 29 % of those with elementary public-school education.

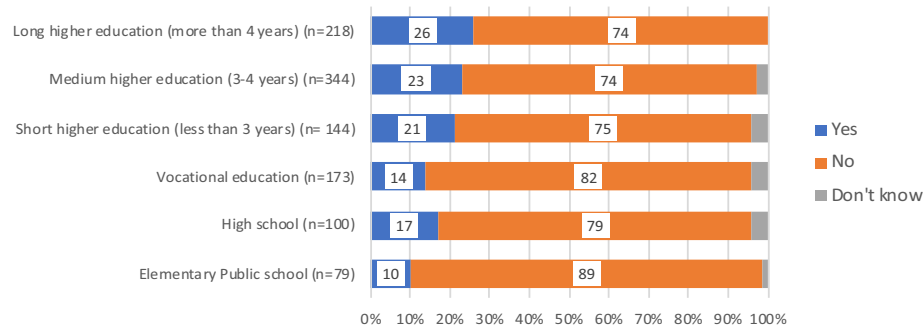


Figure 1. Number of citizens who indicate they have used e-health services to manage their health (e.g. sleep monitoring, smoking cessation, nutrition or exercise).

There are 59% to 64% of citizens who expect that the use of health-IT will improve the quality of the healthcare services they get. Only 10% -12% are concerned that it might impair the quality of health care services. About every four indicate that they have not made up their mind. As shown in figure 2 the attitudes have been quite stable in the period where the studies have taken place.

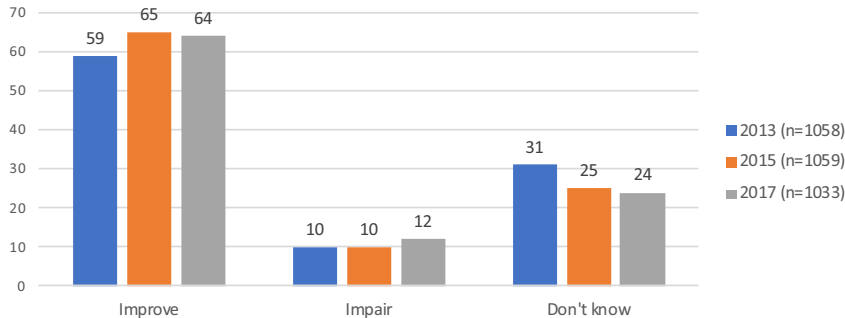


Figure 2. Citizens expectations to whether the use of health-IT will improve or impair the quality of the healthcare services they receive.

In the question of clinicians’ attitude to citizens potential benefits of the national health-IT, almost one third indicate that they do not think the health-IT initiated by the national initiatives will improve the citizens engagement in their own health situation

(Figure 3). Of these just under 8% disagree strongly. On the other hand, more than 20% agree that health-IT can enhance the citizens engagement in health issues, of which only just over 1% strongly agree. More than half of the responding clinicians indicate a neutral attitude, or they don't know.

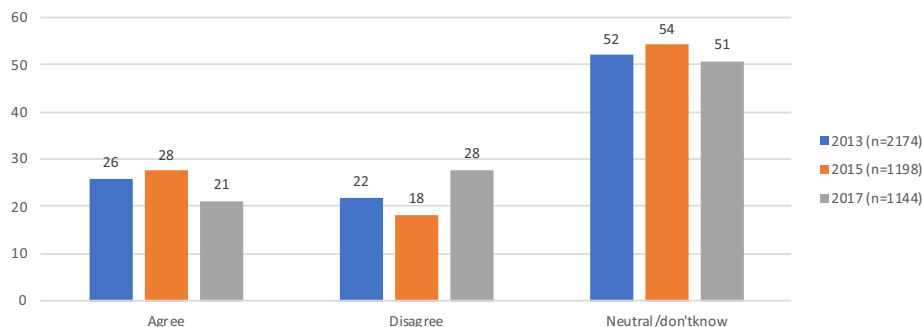


Figure 3. Clinicians attitude to the statement: The national health IT initiatives will strengthen the citizens engagement in their own health

When clinicians are asked about their opinion on how patient can benefit from health-IT by having improved capabilities to master own diseases more than 30% strongly disagree or disagree (Figure 4). Half of the responding clinicians are neutral, or they don't know. Only 18% indicate they strongly agree or agree. This question was not part of surveys before 2015.

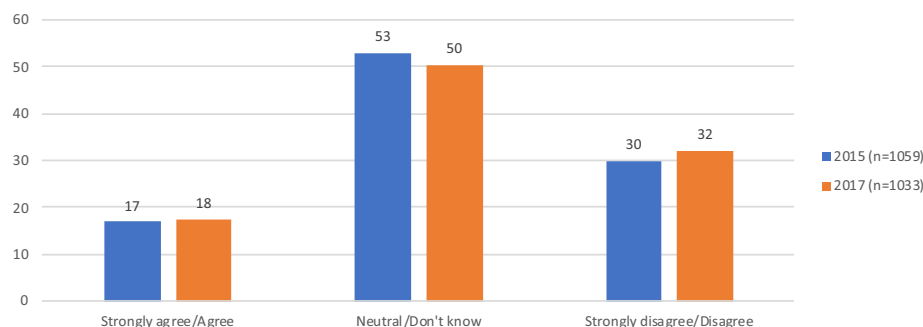


Figure 4. Clinicians response to the statement: I believe that the digital tools we provide to the patients give them better options to master own diseases. (2017, in %, n=1144)

4. Discussion

The objective of this paper was to provide insight into citizens and clinicians perspective on HIT. Despite a well-developed Danish health care service high inequality in health pertains [6]. Our data shows how the citizens with highest education have used e-health services more than twice as much as citizens with the lowest education (26% vs 10%). This, notwithstanding the fact that those citizens in most need of health care service are to be found among those with lowest education [7]. This result is very much in line with the data from the German study [5], that also find a socio-economic difference in the use of HIT, and call for strategic awareness among health care managers and e-health system developers on how to target those citizens in most need of health care when developing

HIT interventions? Awareness of how to beat inequality in use at an early stage when designing health IT is needed. Health inventions need to be context sensitive in its design and need to be supported by health care professionals, to have high impact [8]. We claim this should also be considered for development of HIT. The demographic development, comorbidity, increased expenses should work as enough incentives.

Further, data show that citizens and clinicians have different opinions on the role that HIT can play to enhance citizens engagement and ability to master own disease. Citizens express a far more positive predictions than clinicians as to whether HIT initiatives will strengthen the citizens engagement in their own health or not.

Clinicians also express a dis-satisfaction with the available HIT tools, as to whether they actually do give a better option for the patients to master their health. Whether this view is nourished by clinicians' experience of patient's difficulties in handling HIT tools available today, or from their own disbelieve in HIT tools, are not known but could be further investigated by qualitative research methods. The HIT technology needs to be understood as a socio-technical intervention and therefore, designed and targeted with attention to the context of its use and its users. Finally, we like to draw attention to limitation being the difficulties we have faced in getting a high response rate in the self-selected clinician survey. However, this is not specific for our study but common for this type of studies.

References

- [1] WHO, WHO global strategy on people-centred and integrated health services Interim Report, 2015. www.who.int (accessed October 11, 2019).
- [2] Danish Ministry of Health, Digital Health Strategy 2018-2022, Copenhagen Denmark, 2018. <https://www.healthcaredenmark.dk/media/1611539/The-Danish-Digitalisation-strategy-2018-2022.pdf>.
- [3] M. Almalki, K. Gray, and F.M. Sanchez, The use of self-quantification systems for personal health information: big data management activities and prospects, *Heal. Inf. Sci. Syst.* **3** (2015) S1. doi:10.1186/2047-2501-3-S1-S1.
- [4] J.G. Anderson, M.R. Rainey, and G. Eysenbach, The Impact of CyberHealthcare on the Physician–Patient Relationship, *J. Med. Syst.* **27** (2003) 67–84. doi:10.1023/A:1021061229743.
- [5] C. Ernsting, S.U. Dombrowski, M. Oedekoven, J.L. O'Sullivan, M. Kanzler, A. Kuhlmeier, and P. Gellert, Using Smartphones and Health Apps to Change and Manage Health Behaviors: A Population-Based Survey, *J Med Internet Res.* **19** (2017) e101. doi:10.2196/jmir.6838.
- [6] Diderichsen, Finn, Ingelise Andersen, & Cecilie Manuel (2011). *Ulighed i sundhed – Årsager og indsatser*, Sundhedsstyrelsen, København (report in Danish)
- [7] Dewalt DA, Berkman ND, Sheridan S, et al. Literacy and health outcomes: a systematic review of the literature. *J Gen Intern Med.* 2004;19(12):1228–1239.
- [8] E. Nolte, WHO policy brief. How do we ensure that innovation in health service delivery and organization is implemented, sustained and spread? ISSN 1097-8073
- [9] L. Amo. Education-Based Gaps in eHealth: A Weighted Logistic Regression Approach. *J Med Internet Res.* (2016); 18(10), e267. doi:10.2196/jmir.5188

eSource-Enabled vs. Traditional Clinical Trial Data Collection Methods: A Site-Level Economic Analysis

Eric L. EISENSTEIN,^{a,1} Maryam Y. GARZA^b,
Mitra ROCCA^c, Gideon S. GORDON^c, Meredith ZOZUS^d

^a*Duke University, Durham, NC, United States*

^b*University of Arkansas for Medical Sciences, Little Rock, AR, United States*

^c*United States Food and Drug Administration, Silver Springs, MD, United States*

^d*University of Texas Health Science Center at San Antonio, TX, United States*

Abstract. Directly extracting data from site electronic health records for updating clinical trial databases (eSource) can reduce site data collection times and errors. We conducted a study to determine clinical trial characteristics that make eSource vs. traditional data collection methods more and less economically attractive. The number of patients a site enrolls, the number of study data elements, study coordinator data collection times, and the percent of study data elements that can be extracted via eSource software all impact eSource economic attractiveness. However, these factors may not impact all clinical trial designs in the same way.

Keywords. economic analysis, clinical trial, data collection, eSource, HL7 FHIR

1. Introduction

The complexity of clinical trial protocols continues to increase and is associated with higher clinical trial total and per patient costs [1]. Site-related costs account for 70% of late phase clinical trial total costs [2]. While remote monitoring has been shown to reduce site management costs by reducing the number of on-site monitoring visits [3], there has been less success in reducing costs associated with site personnel workload. Directly extracting data from site electronic health records (EHRs) for updating clinical trial databases (eSource software) has been shown to reduce site data collection time and errors [4,5]. However, previous studies did not consider the costs of the informatics infrastructure needed for sites participating in eSource-enabled clinical trials.

Decision analytic methods are used to simulate complex decisions by combining disparate data sources [6,7]. These tools are particularly useful where there is uncertainty regarding key decision parameters. We conducted a study to determine which clinical trial characteristics make eSource-enabled versus traditional data collection methods more and less economically attractive for clinical trial sites.

¹ Corresponding Author E-mail: eric.eisenstein@duke.edu.

2. Methods

Model Design: We developed a decision analytic model to compare per patient total data collection costs for eSource-enabled vs. traditional data collection methods at clinical trial sites. Our base case model uses characteristics of the TRANSFORM-HF clinical trial [8]. Briefly, TRANSFORM-HF is a 6000 patient, 50-site pragmatic clinical trial to evaluate the use of two loop diuretic medications in patients hospitalized for new or worsening heart failure. Sites in this trial are responsible for collecting information related to the patient’s enrolling hospitalization and a centralized call center conducts follow-up patient interviews.

Decision Model Structure: The decision analytic model used here was constructed to evaluate two data collection strategies: eSource-enabled and traditional (i.e., manual data abstraction from medical records by humans). During the patient’s enrolling hospitalization, three types of data collection occur: (1) a study coordinator collects and enters hospitalization data, (2) the study coordinator resolves queries regarding those data, and (3) an on-site monitor verifies selected cases against the source data. Queries are divided into those that can be resolved while the study coordinator is still in the patient’s EHR and those subsequent queries that require the study coordinator to reopen the patient’s record in the EHR. The study’s model includes four types of measurement data: (1) clinical trial site characteristics, (2) study personnel costs and productivity, (3) eSource costs and productivity, and (4) study outcomes.

Clinical Trial Characteristics: Our model includes six clinical trial site characteristics: (1) number of patients enrolled, (2) number of data elements collected for the study database, (3) percent of study data elements that are FHIR accessible (can be accessed via eSource software using the HL7 Fast Healthcare Interoperability Resources (FHIR) standard) [9], (4) number of queries generated, (5) number of onsite monitoring visits and (6) number of cases reviewed per monitoring visit (Table 1). The TRANSFORM-HF study design assumes that each site will randomize 120 patients on average during a 24-month enrollment period.

Garza et al. identified 155 unique TRANSFORM-HF data elements of which 82 (52.9%) could be mapped to the US Common Core and would be accessible using HL7 FHIR [9]. In a review of 14 protocols for which study data were managed by our institution, queries were generated for 1.5% of study data elements. Based upon operational experience with multicenter studies coordinated by our institution, approximately, 67% of these queries would be resolved during initial data entry while the site coordinator was in the patient’s medical record and 33% would be resolved after initial data entry was completed, requiring the patient’s medical record to be reopened and reexamined. We assumed a risk-based monitoring strategy where each site would have one monitoring visit, 30% of sites would have a second visit and 10% of sites would have a third visit (average 1.4 visits per site). We also assumed monitors would verify source documents for 5 patients per visit (average of 7 cases for all visits).

Table 1. TRANSFORM-HF Clinical Trial Characteristics

Variable	Estimate	Range Tested (+/- 50%)	Data Source
Patients randomized	120	60 to 180	Protocol (Ref 8)
Study data elements	155	78 to 232	Garza (Ref 9)
FHIR accessible rate (%)	52.9%	27% to 79%	Garza (Ref 9)

Study Personnel Costs and Productivity: We used keystroke level modeling to estimate the number of data elements a study coordinator would collect and enter per hour for (1) initial data entry, (2) query resolution during initial data entry while in the patient’s medical record and (3) query resolution after initial data entry requiring the medical record to be reopened and reexamined (Table 2). Keystroke level modeling estimates the length of time an expert will take to accomplish an interactive computer system task without errors [10,11]. We then estimated the number of data elements site personnel would map and confirm to FHIR resources per hour as well as the hourly costs for study site and monitoring personnel. Although site monitors typically are paid by study sponsors (and are not included in site personnel costs), we included monitoring costs to obtain estimates of all data collection costs incurred at clinical trial sites. Hourly costs were estimated by inflating annual salary costs by a 35% fringe benefit rate and 50% indirect costs. Based upon our experience, we assumed personnel could allocate a maximum of 30 hours per week to specific projects.

Table 2. Personnel Characteristics

Variable	Estimate	Source	
TRANSFORM-HF			
Site coordinator data collection rate	134 data elements per hour	Keystroke level model	
Number of queries resolved			
While in patient’s medical record	134 data elements per hour	Keystroke level model	
No longer in patient’s medical record	106 data elements per hour	Keystroke level model	
FHIR Mapping			
Data element mapping	10 data elements per hour	Observation	
Data element mapping time	15.5 hours	Computation	
Data element confirmation	30 data elements per hour	Estimation	
Data element confirmation time	5.2 hours	Computation	
Personnel Costs	Annual	Hourly	
Site principal investigator	\$200,000	\$260	UAMS
Site study coordinator	\$80,000	\$104	UAMS
Information technologist	\$65,000	\$84	UAMS
Information security officer	\$100,000	\$130	UAMS
Site monitor	\$80,000	\$104	UAMS

*UAMS is the University of Arkansas for Medical Sciences

eSource Costs and Productivity: eSource software implementation costs have fixed and variable components. The fixed costs are incurred to set-up a trial and the variable costs are incurred per study data element. Fixed costs include (1) study materials presented to site team members and (2) site eSource security review/approval, setting up the EHR research record, connectivity testing and eSource user training. Variable costs include the time to map study data elements to site EHR FHIR resources and to confirm those mappings. For study material presentation and review, we estimated the site principal investigator, site study coordinator, an information technologist and the site security officer each would require 5 hours (20 total hours, \$2890 cost). For eSource security, we estimated the site coordinator would require 7 hours, the information technologist 4 hours, and the security officer 4 hours (15 total hours, \$1584). Lastly, we estimated the site coordinator and information technologist would require 15.5 hours each to map the TRANSFORM-HF data elements (n=155) to EHR FHIR resources, and the site coordinator would require an additional 5.2 hours to confirm these mappings (36.2 total hours, \$4474 cost) (Table 2).

Study Outcomes and Analyses: The primary study outcome is total per patient data collection costs for each strategy (eSource-enabled and traditional). Secondary outcomes include the major components of total costs and study coordinator costs. The study’s model estimates primary and secondary outcomes and sensitivity analyses. For each data collection strategy, we consider the lower cost strategy to be preferred. Sensitivity analyses investigated the impact of 50% increases and decreases in clinical trial characteristics upon per patient data collection costs. These analyses also identified threshold values for clinical trial characteristics that make eSource-enabled the preferred data collection strategy vs. traditional data collection.

3. Results

Cost Outcomes: The study’s model estimated TRANSFORM-HF per patient total data collection cost was \$10 lower for eSource-enabled versus traditional methods (\$119 vs. \$129) (Table 3). Coordinator costs were \$68 lower (\$61 vs. \$129) while the per-patient cost for eSource software implementation was \$58 higher. Thus, for a site that could enroll 120 patients over the 24-month accrual period, investing in eSource infrastructure would be marginally economically attractive. Coordinator cost savings included \$64 for initial data collection, \$1 query resolution, and \$4 onsite monitoring.

Table 3. TRANSFORM-HF Cost Outcomes

Cost Outcome	Data Collection Method		Difference
	eSource-Enabled	Traditional	
Total Costs	\$119	\$129	(\$10)
Coordinator component	\$61	\$129	(\$68)
eSource component	\$58	\$0	\$58
Coordinator Costs			
Initial data collection	\$57	\$120	(\$64)
Query resolution	\$1	\$2	(\$1)
Site monitoring	\$3	\$7	(\$4)

Sensitivity Analyses: The benefit of eSource-enabled vs. traditional data collection in TRANSFORM-HF clearly was influenced by clinical trial characteristics (Table 4). With a small number of patients, a small study database (number of data elements), a lower FHIR accessible rate (percent study data elements accessible via eSource software), and a higher site coordinator data entry rate, the use of eSource vs. traditional data collection would not be cost saving. However, at higher values for these parameters, the use of eSource software becomes much more economically attractive.

Table 4. TRANSFORM-HF Sensitivity Analysis

Variable	Less 50%	Strategy Cost Difference	Plus 50%	Strategy Cost Difference	Break Even Value	Total Strategy Cost
Number of patients	60	\$45	180	(\$28)	102	\$130
Study data elements	78	\$14	232	(\$33)	123	\$102
FHIR accessible rate	27%	\$35	79%	(\$55)	47%	\$130
Data elements / hour	67	(\$74)	201	\$12	158	\$111

4. Discussion and Conclusion

Our results demonstrate that eSource-enabled vs. traditional data collection methods can be cost saving for the typical TRANSFORM-HF clinical trial site. Our decision analytic model estimated a \$68 per patient reduction in site coordinator data collection costs and a \$58 per patient increase in eSource-related costs. However, these results vary greatly depending upon the number of patients a site enrolls, the number of study data elements and the percent of those data elements that are accessible via HL7 FHIR.

Our study was confined to the economics of sites contemplating the use of eSource software. This work contrasts with the EHR4CR cost-benefit analysis that took the pharmaceutical company perspective [12]. We believe the site perspective is both important and under-investigated. If the use of eSource is not economically attractive for sites, adoption of this technology is unlikely unless clinical trial sponsors include additional study start-up compensation.

If eSource-enabled clinical trials are to become a reality, clinical trial coordinating centers and sites will need to have personnel with the requisite skill sets to implement and manage eSource software. Our model omitted potential eSource benefits such as greater data quality and shorter clinical trial durations. Nonetheless, we believe we have started an important dialogue that could lead to greater recognition of the value of eSource-enabled data collection as we learn more about the ways in which clinical trial design factors drive the economics of eSource-enhanced data collection.

References

- [1] K. Getz, R.A. Campo, New benchmarks characterizing growth in protocol design complexity. *Ther Innov Requi Sci*. 2018;52(1):22-28
- [2] E.L. Eisenstein, P.W. Lemons, B.E. Tardiff et al., Reducing the costs of phase III cardiovascular clinical trials. *Am Heart J*. 2005;149(3):482-8.
- [3] E.L. Eisenstein, R. Collins, B.S Crackness et al., Sensible approaches for reducing clinical trial costs. *Clin Trials*. 2008;5(1):75-84.
- [4] M. Garza, S. Myneni, A. Nordo et al., eSource for standardized health information exchange in clinical research: a systematic review. *Stud Health Technol Inform*. 2019;257:115-124.
- [5] A.M. Nordo, E.L. Eisenstein, J Hawley et al., A comparative effectiveness study of eSource used for data capture for a clinical research registry. *Int J Med Inform* 2017;103:89-94.
- [6] P. Garg, D.J Cohen, T. Gaziano, L. Mauri, Balancing the risks of restenosis and stent thrombosis in bare-metal versus drug-eluting stents. *J Am Coll Cardiol* 2008;51:1844-53.
- [7] B. Levitan, K. Getz, E.L. Eisenstein et al., Assessing the financial value of patient engagement: a quantitative approach from CTTI's Patient Groups and Clinical Trials Project. *Ther Innov Regul Scie*. 2018;52(2):220-229.
- [8] TRANSFORM-HF Protocol. Available at clinicaltrials.gov/ct2show/NCT0326813.
- [9] M.Y. Garza, M. Rutherford, Z. Wang et al., Evaluating the coverage of the HLF FHIR standard to support eSource data exchange implementation for use in multi-site clinical research studies. *AMIA 2020 Informatics Summit*, Houston (Accepted).
- [10] M. Nahm, V.D. Nguyen, E. Razzouk et al., Distributed cognition artifacts on clinical research data collection forms. *Summit Trans Bioinform* 2010.
- [11] D. Kieras, Using the keystroke-level model to estimation execution times. University of Michigan, 2001.
- [12] A. Beresniak, A. Schmidt, J. Proeve et al., Cost-benefit assessment of using electronic health records data for clinical research versus current practices: Contribution of the Electronic Health Records for Clinical Research (EHR4CR) European Project. *Contemp Clin Trials*. 2016;46:85-91.

Health Professionals' Experiences of the Benefits and Challenges of Online Symptom Checkers

Sari KUJALA^{a,1}, Iris HÖRHAMMER^a, Riitta HÄNNINEN-ERVASTI^b,
Tarja HEPONIEMI^c

^a*Aalto University, Finland*

^b*University of Oulu, Finland*

^c*Finnish Institute for Health and Welfare, Finland*

Abstract. Online symptom checkers and assessment services are used by patients seeking guidance on health problems. In this study, the goal was to identify health professionals' experiences of the benefits and challenges of new symptom checkers providing triage advice. Data was collected through an online survey of 61 health professionals who were target users of the online symptom checkers implemented in six public health organizations and one private occupational health clinic. Most of the health professionals supported the use of online symptom checkers and found services useful to patients because they provided patients quick contact with health professionals and referral to care or self-management instructions regardless of time and place. Health professionals were less confident that most of the patients were capable and willing to use the symptom checkers. Health professionals were satisfied with symptom checkers providing them with more useful information before meeting patients. By contrast, symptom checkers were seen as disrupting clinical work and time-consuming. The results imply that the clinical work processes should be redesigned to guide patients in an efficient manner, avoid work overlap, and provide work motivation for professionals.

Keywords. Online symptom checkers, triage advice, health professionals

1. Introduction

Online symptom checkers and assessment services are used by patients seeking guidance on health problems [1]. Using computerized algorithms, symptom checkers ask patients questions about their symptoms, provide them with a potential diagnosis, suggest self-care, and direct them to the appropriate care setting [2]. The goal is often to supplement or replace resource-intensive telephone triage lines. The online symptom checkers also have the potential to support self-care and decrease the number of unnecessary visits.

Based on Chambers' et al. [1] recent literature review, the evidence of the diagnostic accuracy and impacts of online symptom checkers remains weak. Triage advice provided by symptom checker algorithms tended to be more risk-averse than that given by health professionals. Thus, symptom checkers may encourage patients to seek care in cases where self-care would be more justified [2]. In addition, younger and more highly-educated people are more likely to use online symptom checkers, whereas older and less

¹ Sari Kujala, Corresponding author, E-mail: firstname.secondname@aalto.fi.

educated patients are more likely to prefer telephone or face-to-face contact [1]. Thus, health professionals through their endorsement play an important role in supporting patients and increasing their trust in technical solutions [3]. If only younger and more highly educated patients use online symptom checkers, the cost-effectiveness goals of these services may not be reached, and health equity issues may arise.

In Finland, a joint project of 14 of the largest cities and hospital districts have been developing a patient portal called Omaolo. The three main services include an online symptom checker called Oirearvio, a patient self-assessment and well-being coaching, and a health care plan. In spring 2017, health professionals' pre-implementation expectations relating to the patient portal were studied through an online survey [4]. Most of the 2,943 respondents (74%) expected benefits for patients; however, many expressed concerns that patients were unwilling or unable to use the new service. As symptom checkers are a new kind of health technology and the personnel appear to have concerns, it is important to follow-up on their implementation. Follow-up activities were also found to be generally important to support personnel during change, avoid resistance, and remove barriers [5].

In 2018, healthcare organizations started to use online symptom checkers, which were the first finalized functionality of the patient portal. The symptom checkers do not provide diagnosis, but they are used to direct the patient to obtain appropriate care or provide self-care instructions. The symptom checkers were designed for specific conditions: low back pain, urinary tract infection, and upper respiratory tract infection. The goal of this survey study was to identify health professionals' first experiences of using symptom checkers. Findings provide a better understanding of the role of symptom checkers in clinical work and the views of health professionals' relating to the benefits and challenges of the system.

2. Methods

An online questionnaire was developed to identify the experiences of health professionals, including nurses, physiotherapists, and physicians. The questionnaire was a new version of the previous expectation questionnaire including both multiple-choice and open-ended questions [4]. Health professionals were asked to evaluate their current experiences (instead of expectations as in the previous questionnaire), and the wording was changed from future to present tense.

The questionnaire included existing validated survey items for measuring professionals' support for the symptom checkers [6], their usability [7], and their influence on professional autonomy [8]. In addition, the participants were asked to rate whether the symptom checkers had brought the planned benefits to their work and patients and whether good implementation practices were used in their organization. The scales ranged from 1 (fully disagree) to 5 (fully agree) and included a sixth option, 6 (I do not know), which was removed from the analysis. Two open-ended questions were related to the benefits and challenges of the symptom checkers. In the background questions, participants were also asked how often they had used the symptom checkers. The quantitative data were with using descriptive statistics, and the responses to open-ended questions were content-analyzed.

The data was gathered from December 2018 to March 2019. The survey invitation with the link to the questionnaire was sent to health professionals identified by the local project managers as potential users of the symptom checkers. The respondents

represented six public health organizations and one private occupational health clinic that was located in the largest cities of Finland. The seven organizations were selected for the study as they were the first ones to adopt the symptom checkers. Although participation was anonymous, respondents could submit their email address at the end of the survey to participate in a draw for 10 movie tickets.

3. Results

A total of 61 health professionals answered the questionnaire; the response rate was 15.2% in six of the organizations (unfortunately, we were unable to identify a response rate for one organization). The mean age of the respondents was 43.8 years (SD = 10.4), and 87% were female. Most of the respondents (77%) were nurses, 15% were physiotherapists, and 7% physicians. The use of the symptom checkers was not yet very established: Most of the respondents (41%) had used them one to two times per month, 39% weekly or daily. Furthermore, 18% had only tried, and one respondent (1.6%) had not used them at all. A few of the respondents also commented that the symptom checkers were not yet widely used among patients.

Most of the health professionals (88.5%) agreed that they supported the use of online health checkers. The majority of them also agreed that the symptom checkers resulted in planned patient benefits (Table 1). In particular, respondents believed that the symptom checkers improved the availability of the services. However, the respondents were clearly less confident that most of the patients were capable and willing to use the symptom checkers.

Table 1. Percentage of health professionals agreeing with the statements related to the patient benefits of the symptom checkers.

Patient benefit	Professionals agree (%)
The symptom checker improves the availability of the services.	79
By using the symptom checker, patients have the same criteria to receive care.	68
The symptom checker supports individual care.	62
I believe that over half of the patients are capable of using the symptom checker.	50
I believe that over half of patients are willing to use the symptom checker.	42

Table 2. Percentage of health professionals agreeing with the statements related to the benefits of the symptom checkers for their work.

Benefit for professionals' work	Professionals agree (%)
Using the symptom checker provides me with more useful information before meeting a patient.	67
The symptom checker leaves more time for meeting with patients.	45
The symptom checker reduces the amount of routine work in my tasks.	38
The symptom checker reduces my work.	33
The symptom checker decreases my control over the patient process.	16
The symptom checker decreases my control over professional decisions.	12

Most health professionals also agreed that the symptom checkers were useful for their work because it provided more useful information about patients before meeting them (Table 2). However, less than half of them believed that the symptom checkers

would be efficient in their work. A minority also believed that symptom checkers decreased their control over the patient portal and professional decisions.

In open-question responses, the most often mentioned benefits included the following: a) patients will be guided to the right professional (a nurse, physiotherapist or doctor), b) patients get quick help and standardized self-management instructions, c) there will be less unnecessary visits, fewer phone calls, and less queuing. The most often mentioned challenges were related to the changes that the new symptom checkers brought to work, responsibilities, and common work processes. The symptom checkers were seen as added work, as they created one more channel to handle, and some respondents mentioned that patients started to unnecessarily use several contact methods. Many were also concerned that not all patients were able to use computers or the symptom checkers and believed that the care instructions should be made easier so patients could understand. For example, one respondent mentioned that the symptom checkers required more work than a phone call because patients did not know how to use them. Furthermore, the symptom checkers were seen as needing development, as they did not notify nurses when a patient made contact, and they also required extra work as nurses needed to transfer patient health information to a separate system.

4. Conclusions

The majority of the health professionals had a positive view of the new online symptom checkers. Most of them accepted the use of online symptom checkers, and they found that the services were useful to patients because they provided patients with a method of quick contact with health professionals and referral to care or self-management instructions, regardless of time and place. Health professionals valued the fact that the symptom checkers provided them with more patient information before meeting patients. Some of the professionals also saw opportunities to save resources as patients were immediately guided to see the right professional and, in some cases, patients could manage the health issues by themselves.

On the other hand, the symptom checkers clearly challenged existing work practices and the roles of health professionals. Health professionals believed that symptom checkers created extra work for them, considering that the services offered patients a new channel to contact health professionals. Some of the health professionals believed that the symptom checkers were a threat to their professional autonomy as they reduced their control over the patient process and professional decision making. Previously, this perceived threat was found to have a significant negative influence on health professionals' acceptance of new technology [4,8]. Furthermore, in our studied organizations, a minority of the health professionals resisted the new symptom checkers. Thus, there is a risk that teams would not work optimally and some professionals would not recommend the symptom checkers to patients.

The results imply that the implementation of symptom checkers, new work processes, and division of work require special attention in health organizations. Health professionals need support and training for this novel situation, and work processes should be redesigned to guide patients efficiently, avoid work overlap, and ensure work motivation among health professionals. The technical solutions should support health professionals' work in a holistic manner, so that they are not separated from other systems used and that incoming messages notify the responsible person.

Successful implementation of symptom checkers also requires patients to be engaged and motivated to use them. The variety of patients and their skills makes this challenging. Respondents indicated that patients lacked skills to use computers and symptom checkers. Although the content of the symptom checkers was built on scientific evidence, it is possible that they could lead to wrong conclusions if patients misinterpret the questions used. Furthermore, one respondent commented that the symptom checkers appeared to create concerns and lack of support among patients. Thus, it is important that symptom checkers are easy to understand and use. In addition, patients' experiences should be studied to identify their support needs and barriers of use.

The sample size of this survey study was relatively small as the online symptom checkers were used less than one year in a few organizations. This study focused on health professionals' subjective experiences in the early phase of the adoption process. As health professionals and patients will gain more experience in using the online symptom checkers, usage may become more routinized and smoother. However, providing a better understanding of the health professionals' views of symptom checkers, and their benefits and challenges will help developers in designing technical solutions and health organizations in supporting the adoption of the new services.

Acknowledgements

The authors are grateful to the participating health organizations and project managers. This work was supported by the Strategic Research Council at the Academy of Finland, decision numbers 303606, 327145 and 327147.

References

- [1] D. Chambers, A.J. Cantrell, M. Johnson, L. Preston, S.K. Baxter, A. Booth, J. Turner, Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review, *BMJ Open*. 9 (2019) e027743. doi:10.1136/bmjopen-2018-027743.
- [2] H.L. Semigran, J.A. Linder, C. Gidengil, A. Mehrotra, Evaluation of symptom checkers for self diagnosis and triage: audit study, *BMJ*. (2015) h3480. doi:10.1136/bmj.h3480.
- [3] S. O'Connor, P. Hanlon, C.A. O'Donnell, S. Garcia, J. Glanville, F.S. Mair, Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies, *BMC Med Inform Decis Mak*. 16 (2016) 120. doi:10.1186/s12911-016-0359-3.
- [4] S. Kujala, I. Hörhammer, J. Kaipio, T. Heponiemi, Health professionals' expectations of a national patient portal for self-management, *International Journal of Medical Informatics*. 117 (2018) 82–87. doi:10.1016/j.ijmedinf.2018.06.005.
- [5] T. Ingebrigtsen, A. Georgiou, R. Clay-Williams, F. Magrabi, A. Hordern, M. Prgommet, J. Li, J. Westbrook, J. Braithwaite, The impact of clinical leadership on health information technology adoption: Systematic review, *International Journal of Medical Informatics*. 83 (2014) 393–405. doi:10.1016/j.ijmedinf.2014.02.005.
- [7] R.L. Wakefield, D. Whitten, Mobile computing: a user study on hedonic/utilitarian mobile device usage, *European Journal of Information Systems*. 15 (2006) 292–300. doi:10.1057/palgrave.ejis.3000619.
- [8] K. Finstad, The usability metric for user experience, *Interacting with Computers*. 22 (2010) 323–327.
- [9] Z. Walter, M.S. Lopez, Physician acceptance of information technologies: Role of perceived threat to professional autonomy, *Decision Support Systems*. 46 (2008) 206–215. doi:10.1016/j.dss.2008.06.004.

How Come Nothing Changed? Reflections on the Fasting-Time Project

Morten HERTZUM^{a,1} and Jesper SIMONSEN^b

^aUniversity of Copenhagen, Copenhagen, Denmark

^bRoskilde University, Roskilde, Denmark

Abstract. Many healthcare interventions fail to produce the intended effect. In this paper we look back at the fasting-time project, which aimed to shorten patients' preoperative fasting times. However, the project failed to achieve this effect, even though it had been identified and prioritized by the clinicians at the studied hospital. A set of mutually reinforcing factors collectively explain why the project failed to produce change. The four main factors are: lack of urgency, risk aversion, day-to-day busyness, and lack of managerial commitment at the department level. The simultaneous presence of these factors complicates efforts to counter them.

Keywords. effects realization, implementation, change management, fasting time

1. Introduction

It is nontrivial to improve hospital practices by introducing new information systems and new ways of working. Diffusion-of-innovations theory states that for an innovation to be adopted it must, among other things, be compatible, avoid complexity, provide relative advantage, and be actively championed by people such as opinion leaders [1]. Technology-acceptance research finds that for systems to be adopted and used they must be perceived as useful, easy to use, and even enjoyable [2]. Reviews of the implementation of electronic health records confirm many of these factors and add factors such as adaptability, data migration, implementation climate, management, organizational readiness, planning, staff training, and external policy and incentives [3][4]. In this study, we look back at a project that failed to produce the intended effect, even though the clinicians had themselves identified and prioritized this effect.

The project [5] concerned the reduction of patients' preoperative fasting times. To reduce the risk of pulmonary aspiration due to vomiting during anesthesia, patients must fast for six hours prior to surgical operations. Fasting times in excess of six hours should, if possible, be avoided to reduce postoperative complications and patient discomfort. But avoiding fasting times in excess of six hours presupposes efficient coordination among the involved clinicians. Organizationally, the fasting-time project was to introduce new procedures for recording fasting times and acting on long fasting times. Technologically, the project was to utilize a network of electronic whiteboards for recording, and visualizing, how long the individual patient had fasted. In the following, we first summarize the method and results of the project, then we reflect on why the project failed to produce shorter preoperative fasting times.

¹ Corresponding author; hertzum@hum.ku.dk.

2. Method

The effort to improve the coordination of operations, and thereby reduce fasting times, was a participatory-design project conducted by the authors and a group of staff from a hospital in Denmark. The project followed an effects-driven approach [6]. That is, the project consisted of specifying the purpose of the project in terms of an effect, realizing this effect through technical and organizational changes, assessing the extent to which the specified effect had been achieved, and iterating to achieve a fuller realization of the effect. Table 1 gives the timeline of our empirical activities.

We started with four workshops (a total of 10 hours) during which clinicians from different departments and staff groups at the hospital specified and prioritized the effects to be achieved in the project. The effect of reducing preoperative fasting times received top priority. After these workshops, a small group was established. It consisted of the authors, three staff members (a nurse from the operating ward and a secretary from each of the two surgical departments), and a research assistant. This group met for 16 participatory-design meetings (35 hours) to realize and assess the fasting-time effect. In between the meetings the three staff members worked to implement the group’s decisions in their departments. The group meetings were supplemented with observation (70 hours) of how the operations were coordinated at the hospital. During the effects assessment, fasting times were recorded and visualized for three months.

Table 1. Timeline of the empirical activities, which spanned the period September 2014 – December 2015

Date	Activity
<i>Phase 1: effects specification</i>	
Sep 18	Workshop with 5 clinicians and a hospital IT project manager to specify effects
Sep 26	Workshop with 10 clinicians and a hospital IT project manager to specify effects
Nov 7	Workshop with 7 clinicians to specify effects
Dec 12	Workshop with 9 clinicians to prioritize and elaborate effects
<i>Phase 2: effects realization</i>	
Feb 17	Group meeting to kick off the realization of the fasting-time effect
Feb 20	Group meeting to plan the project activities and begin defining fasting time in detail
Feb 26	Observation at surgical departments to get a sense of the coordination of operations
Feb 27	Observation at operating ward to get a sense of the coordination of operations
Mar 6	Group meeting to define fasting time (and how to record it) in detail
Mar 17	Workshop with whiteboard vendor to configure the fasting-time fields
Mar 27	Group meeting to devise a standard procedure for the recording of fasting times
Apr 10	Group meeting to ensure the adoption of the whiteboard fields and standard procedure
Apr 24	Group meeting to promote the project and align it with the current use of the whiteboard
May 8	Group meeting to finalize the setup of the effects assessment: the fasting-time recordings
May 22	Group meeting to organize and schedule the follow-up on the fasting-time recordings
<i>Phase 3: effects assessment</i>	
Jun 4	Group meeting to prepare the next phases and the analysis of the fasting-time recordings
May 11 - Aug 14	Fasting times recorded and visualized on the whiteboard
May 18 - Jun 30	Observation at the surgical departments of how fasting times were recorded and used
Aug 21	Group meeting to validate the fasting-time recordings and brainstorm new interventions
Sep 4	Group meeting to discuss the fasting-time recordings with department management
<i>Phase 4: effects realization</i>	
Sep 13 - Oct 4	Observation of whiteboard meetings at the operating ward (four Sundays)
Sep 18	Group meeting to analyze project progress and identify barriers that curbed progress
Oct 2	Group meeting to prepare a project presentation and discuss the whiteboard meetings
Oct 23	Group meeting to analyze project progress and identify barriers that curbed progress
Nov 5	Group meeting to devise a renewed intervention: daily whiteboard meetings
Nov 16 - Dec 15	Observation of whiteboard meetings at the operating ward (daily for a month)
Dec 11	Group meeting to discuss lessons learned by the project group and the hospital at large

3. Results

In brief, the changes made to reduce the patients' fasting times consisted of (a) defining fasting time in meticulous detail, (b) extending the hospital-wide network of electronic whiteboards with fields showing each patient's fasting time, (c) devising a standard procedure for the recording of the fasting time on the whiteboard, (d) instituting this procedure among the clinicians, (e) bringing fasting times to the attention of staff and management, (f) adjusting the instructions given to patients about when to eat for the last time before arriving for elective operations, and (g) introducing daily whiteboard meetings at the operating ward. During the daily whiteboard meetings, the surgeons reviewed the patients scheduled for operation, fine-tuned the schedule to accommodate acute operations and reduce fasting times, and had the possibility to order a meal for patients who would not be operated within the following six hours. For more detail about the changes made to reduce the fasting times, see Simonsen et al. [5].

In the period May 11 – August 14, 2015, the fasting time was recorded for 416 (32%) of the patients who were operated upon during this period of time. The average fasting time was 12.8 hours, more than twice the required 6 hours. The top 10% of the fasting times were 18-23 hours for acute patients and 16-20 hours for elective patients. In assessing the fasting times, a chief physician expressed surprise that they were so long, especially for the acute patients. Another chief physician agreed that the fasting times were long and likely to cause postoperative complications, such as prolonged wound healing. He also asked the pertinent question: "*Who assumes responsibility for this?*" His own tentative answer was that on a daily basis no one really cared about long fasting times. Our observations of the daily whiteboard meetings, introduced to heighten awareness of fasting times, confirmed this answer: Many clinicians did not attend these meetings, and on no occasion did we observe that a meal was ordered for a patient who would not be operated within the following six hours. The fasting-time project was discontinued in February 2016 without obtaining shorter fasting times.

4. Discussion

Looking back at the project we contend that no single factor explains the failure to shorten the patients' preoperative fasting times. Rather, multiple factors interact and reinforce each other. The review by Ross et al. [4] makes a similar point, while that by Priestman et al. [3] merely catalogs single factors. The simultaneous presence of interacting factors complicates any attempt to counter the factors and make change happen. In retrospect, we identify four primary and several supplementary factors that collectively explain the outcome of the fasting-time project.

First, *the physicians did not experience an urgent need*. During the effects specification the fasting-time effect was championed by one emergency department (ED) physician in particular. This ED physician saw shortened fasting times as an indicator of improved preoperative coordination. Furthermore, this physician knew the whiteboard technology, which had been in use in the ED for three years prior to its introduction throughout the hospital. In the ED the whiteboard had spawned multiple process innovations. However, the ED physician did not become a member of the fasting-time project group, which throughout the effects realization and assessment struggled to obtain buy-in from the physicians. Shortened fasting times never became an urgent issue for the physicians for two reasons. First, it did not in itself present a

problem they perceived as a threat to their competent performance of their work. In spite of clinical evidence that prolonged fasting times cause postoperative complications, fasting times remained a matter of discomfort to the patients. Second, shortened fasting times did not tie in with a performance indicator the physicians were keen to achieve (see below). Improved preoperative coordination was not recognized as an important goal, but rather seen as a means. Using the improved coordination as a means to attract more patients (who would value waiting less) was discussed favorably by hospital management, but was not a goal that motivated the individual physicians. While lack of urgency is a recognized barrier to change [7], it is not easily dismantled.

Second, *the surgical department was not prepared to risk idle operation slots*. An idle operation slot is the situation where an operating room is ready and staffed with a surgical team but there is no patient to operate upon. One of the key performance indicators at the surgical department was to avoid idle operation slots. To achieve this goal, the operations were scheduled so that there would always be two patients ready for operation. If it turned out that the first patient could for some reason not be operated anyway, then the second could be operated instead. While this practice virtually prevented idle operation slots, it prolonged fasting times: In most cases, the first patient could be operated upon; the second was not operated until later and faced prolonged fasting as a consequence. The practice of always having two patients ready for operation was one of the up-front reasons for believing that changes in how operations were scheduled could realistically lead to shorter fasting times. However, alternative practices that shortened the fasting times increased the risk of occasionally not having any surgical patient who had fasted for the required six hours. Although this risk was small it was influential because it involved a key performance indicator. The aversion of this specific risk may be a conscious priority – filled operation slots over shortened fasting times. However, a more general risk aversion stifles any change [8] because change, by definition, upsets the status quo and thereby incurs uncertainty and risk.

Third, *the clinicians' daily schedules left few resources for change efforts*. The clinicians were busy seeing patients, ordering tests, looking up procedures, keeping records, interpreting test results, consulting colleagues, obtaining equipment, and so forth. As a somewhat extreme example, one of the surgical chief physicians routinely had parallel calendar appointments. These parallel appointments partly indicated his stressful work situation and partly increased the stress on his colleagues, who could never know which appointments he would attend. A large surplus of time for change efforts would have been inappropriate because it would have meant that the hospital was not run in a cost-effective manner. But the current state of affairs was also reproachable. The constant resource and workflow optimizations in Danish healthcare had resulted in the near absence of slack resources at the hospital. The clinicians tended to feel that all their resources were tied up in their performance of their day-to-day activities. Few had the resources necessary to engage in change efforts, such as the new practices intended to shorten fasting times. To the hospital, this meant that it was at risk of going solid [9], that is, of becoming more or less incapable of change.

Fourth, *managerial commitment was lacking at the department level*. The fasting-time project was supported by hospital top management, which advocated the project and its use of the whiteboard. Top management also met with management at the department level to obtain their commitment to the project and negotiate the terms of their participation. However, the departments enjoyed a high degree of discretion, which is common in hospitals [8]. In practice, top management could not order the departments to participate in the project, but merely ask them to do so. While the

department managers accepted the project, they remained uncommitted to it, partly due to its reliance on the whiteboards. The decision to use the network of whiteboards for recording, and visualizing, fasting time was based on the successful use of the whiteboard for such purposes in the ED. In the other departments, the whiteboard had not been similarly successful. For these departments committing to the fasting-time project would also mean committing to turning the whiteboard into a technology that was in regular use by the staff, who was already skeptical toward the whiteboard [10]. While hospital top management liked the prospect of increasing the adoption of the whiteboard through the fasting-time project, it eventually had to accept that the department managers were not going to put their weight behind the project.

In addition to the main factors discussed above, several supplementary factors also contributed to the outcome of the project. These included (a) competing priorities such as the preparations for the upcoming introduction of a new electronic health record across the hospital, (b) difficulty establishing an occasion in the planning of the operations where information about fasting times could influence decisions, (c) the limited interest in information technology among many of the clinicians, (d) the interdepartmental character of the project, which increased its organizational complexity compared to the successful use of the whiteboard within the ED, and (e) the ripple effect of incomplete fasting-time recordings on the subsequent use of these recordings in scheduling the operations, for example at the daily whiteboard meetings.

5. Conclusion

In spite of top management support, considerable participatory-design work, and good intentions, no one really assumed responsibility for shortening patients' preoperative fasting times. The identified set of mutually reinforcing factors collectively explain why nothing changed. The four main reasons are lack of urgency, risk aversion, day-to-day busyness, and lack of managerial commitment at the department level.

References

- [1] E.M. Rogers, *Diffusion of Innovations. Fifth Edition*, Free Press, New York, 2003.
- [2] K. Hornbæk, M. Hertzum, Technology acceptance and user experience: A review of the experiential component in HCI, *ACM Transactions on Computer-Human Interaction* **24**, 5 (2017), article 33.
- [3] W. Priestman, S. Sridharan, H. Vigne, R. Collins, L. Seamer, N.J. Sebire, What to expect from electronic patient record system implementation: Lessons from published evidence, *Journal of Innovation in Health Informatics* **25**, 2 (2018), 92-104.
- [4] J. Ross, F. Stevenson, R. Lau, E. Murray, Factors that influence the implementation of e-health: A systematic review of systematic reviews (an update), *Implementation Science* **11** (2016), article 146.
- [5] J. Simonsen, H. Karasti, M. Hertzum, Infrastructuring and participatory design: Exploring infrastructural inversion as analytic, empirical, and generative, *Computer-Supported Cooperative Work* (in press).
- [6] M. Hertzum, J. Simonsen, Effects-driven IT development: Specifying, realizing, and assessing usage effects, *Scandinavian Journal of Information Systems* **23**, 1 (2011), 3-28.
- [7] J.P. Kotter, *A Sense of Urgency*, Harvard Business Press, Boston, MA, 2008.
- [8] I.M. Nembhard, J.A. Alexander, T.J. Hoff, R. Ramanujam, Why does the quality of health care continue to lag? Insights from management research, *Academy of Management Perspectives* **23**, 1 (2009), 24-42.
- [9] R. Cook, J. Rasmussen, "Going solid": A model of system dynamics and consequences for patient safety, *BMJ Quality & Safety* **14**, 2 (2005), 130-134.
- [10] A. Torkilshegygi, M. Hertzum, Incomplete by design: A study of a design-in-use approach to systems implementation, *Scandinavian Journal of Information Systems* **29**, 2 (2017), article 2.

How to Evaluate Health Applications with Conversational User Interface?

Kerstin DENECKE^{a,1} and Jim WARREN^b

^a Bern University of Applied Sciences, Bern, Switzerland

^b University of Auckland, Auckland, New Zealand

Abstract. Application of conversational user interfaces (CUI) or chatbots to healthcare is gaining interest fueled by the rising power of artificial intelligence, increasing popularity of mobile health applications and the desire for engagement and usability. While their use is mainly justified by increasing adherence to mobile health applications and facilitating interactions with the system, the question arises: How can such systems be evaluated in a reliable manner? This paper introduces an evaluation framework for health systems whose core interaction principle is a CUI. We derive quality dimensions and attributes by collecting relevant evaluation aspects from applications that have been developed in previous work and from literature on health chatbots. The collected aspects are aggregated into six thematic categories for chatbot quality, including user experience, linguistic, task-oriented and artificial intelligence perspectives, but also healthcare quality and system quality perspectives. The framework is intended to support developers and researchers in the domain of chatbots in healthcare in selecting relevant quality attributes to be assessed before their systems are distributed to patients.

Keywords. Evaluation, Health application, Chatbot, Conversational user interface, Chatbot evaluation, Natural language understanding

1. Introduction

Mobile health applications are increasingly used by patients to collect health data, and in this way to continuously monitor personal health and to get support from a virtual personal health coach throughout the day. To realize virtual health coaches or to provide mobile health interventions, conversational user interfaces (CUI) have been gaining in interest with mobile health application developers in recent years [1]. A CUI-based system is a computer program that interacts with users using natural language (written or spoken). The aim of such a system is to simulate a human conversation. To reduce system complexity, the user input is often restricted to selecting specific predefined items (e.g. choosing options as replies). A minority of CUI-based systems allow unconstrained natural language input. Some systems use embodied avatars, while others reduce the conversation to an exchange of text messages. Among healthcare chatbots, we can recognize different application areas or scopes of use. Therapeutic or counselling chatbots provide some specific therapy such as cognitive behavior therapy (CBT) [2]. Disease or medication management chatbots support the user in managing medications, provide knowledge on medication or a disease, remind on intake, explain interactions or

¹ Corresponding Author, Kerstin Denecke, Bern University of Applied Sciences, Quellgasse 21, 2501 Biel, Switzerland; E-mail: kerstin.denecke@bfh.ch.

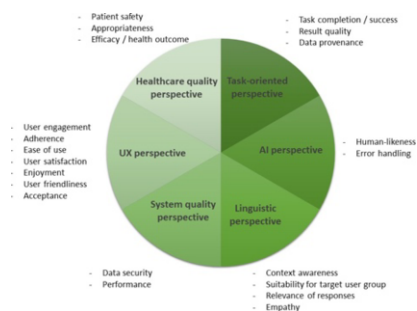
contraindications etc. (e.g. eMMA [3]). Educational applications can be stand-alone or integrated in an application with additional scope of use (e.g. psychoeducation is integrated in a CBT chatbot). Other chatbots are used for screening or collecting the medical history (e.g. Ana [4]). Several applications exist that collect symptoms from a user to make a triage. Finally, chatbots could help in retrieving information such as enabling physicians in getting the relevant information from the electronic health record of a specific patient. Depending on the scope of use of a chatbot, different quality attributes gain in importance when judging the quality. While thousands of health chatbots can be downloaded from app stores, evaluation results are reported only rarely [1]. If at all, only selected aspects are evaluated with results often not comparable due to different methodologies, evaluation aspects and metrics used [1]. In this paper, we address the question: What are the relevant aspects for evaluating healthcare chatbots? The main objective of this work is to provide guidance on relevant dimensions for evaluating health chatbots.

2. Methodology

We are focusing on mobile health applications for patients with CUI that allow unrestricted natural language user input. In previous work, we developed four such applications: eMMA – a medication management assistant [3], CLAIRE – a mobile application with virtual reality and voice user interface to educate patients [5], Ana – a system that collects the music biography as a starting point for music therapy [4] and SERMO – a mental health chatbot that helps in regulating emotions. In previous work, these chatbots have been evaluated using different methods and from various perspectives. However, their clinical effectiveness has not yet been studied. From those works, we collect evaluation aspects. Additionally, we reviewed articles on healthcare chatbots retrieved by a PubMed search with keywords "chatbot, conversational user interface, chatbot, CUI". According to a snowball approach, we added additional works starting from the retrieved results. To develop the framework, we grouped the attributes reported on healthcare chatbots and derived evaluation dimensions.

3. Results

Six perspectives emerge as important for our framework, including user experience, linguistic, task-oriented and artificial intelligence perspectives, as well as healthcare quality and system quality perspectives (Fig. 1).



3.1. Task-oriented perspective

The task-oriented perspective of a chatbot evaluation assesses the capabilities of a chatbot

Figure 1 Evaluation framework. Quality attributes fall into six thematic categories

to perform a specific task such as retrieving information, collecting specific information from a patient or predicting a diagnosis etc. It considers the degree of *task completion* or *task success* and the *result quality*. For example, a chatbot that is supposed to make a triage based on the symptoms described by a user should be assessed with respect to 1) the completeness of collected data relevant for triage and 2) the quality of the classification based on the symptoms. As such, the underlying algorithm should minimize triage classification errors and an appropriate statistic should be selected and measured (e.g. as accuracy, if the triage classes are relatively equal in frequency and in harm associated with confusion of one class for another). As a further requirement to avoid patient harm, the underlying knowledge base of a healthcare chatbot has to be evidence-based and relevant. This issue is included in the framework by the aspect *data provenance* within the task-oriented perspective. An underlying knowledge base also has to be complete with respect to the task that a chatbot is supposed to do.

3.2. Artificial intelligence (AI) perspective

The AI perspective studies to what extent the chatbot is capable of acting like a human being, e.g. in terms of problem solving or influencing a user as well as the dialogue efficiency (Note obviously, this isn't the full scope of AI, but we use 'AI perspective' as a convenient label for this focus on achieving human-like/anthropomorphic behavior). In principle, human-likeness can be studied with the Turing test. Since the Turing test cannot measure emotional engagement with users, the metric conversation turns per session has been introduced for success of social chatbots [6]. However, in healthcare applications, it is important that users are not deceived into over-reliance on a chatbot and led to failure in recognizing its limitations due to its human-like dialog. To address this issue, it could be assessed whether the chatbot implements mechanisms to determine its own intellectual limits, so that it can forward a patient to a human healthcare provider to avoid patient harm. Error handling is another aspect of relevance for healthcare chatbots within the AI perspective. It concerns the capability of a bot to react on unexpected user input or even missing data points.

3.3. System quality perspective

System quality in healthcare chatbots should consider data security and performance (e.g. answering time). Unlike patient-doctor encounters, where patient privacy and confidentiality are protected, healthcare chatbots often do not yet consider these aspects. Some of these systems even run on social media platforms such as Facebook messenger where the use of collected data is unknown to the user or captured in data policies that are long and difficult to understand and to assess. This means the data could be sold, traded or marketed by the distributor of a chatbot. To address this issue, we included data security into the framework. In existing evaluations, aspects related to system quality have not yet been reported. They include among other things that privacy policy must be provided that is application-specific and easily accessible, ideally within the application; and the treatment of confidential data of the chatbot user must be described in detail. Some of the relevant aspects can be assessed using a checklist such as to which third parties data are transmitted, and where data are stored. Further, health chatbots must comply with existing regulations such as the General Data Protection Regulation and Medical Device Regulation.

3.4. Linguistic perspective

Evaluation of a chatbot from the linguistic perspective concerns the effectiveness of a conversation. Relevant aspects include response relevance [7], context-awareness or overall dialogue quality given divergent user input. The latter addresses the fact that the language of bot responses should be suitable for the target user group, i.e. a chatbot designed for interacting with children must use different language than one for adults. Patient-doctor encounters are ideally characterized by empathy, particularly for counseling applications. Thus, a health chatbot should express *empathy* in its dialog, and this is an aspect that should be evaluated for example by adopting the interpersonal communication competence scale [8].

3.5. UX Perspective

The UX perspective evaluates an application from a human factors or usability point of view, i.e. the feasibility of a chatbot solving specific tasks. It is mainly reflecting usability issues, for example adherence, user friendliness, ease of use, appropriateness [4], user engagement [2], user satisfaction, enjoyment. The indicator *acceptance* is a new aspect that has been added to our framework and could not be found in existing healthcare chatbot evaluations. For example, acceptance could be assessed in terms of a classic technology acceptance model (TAM) [9] which combines a user's perception of ease of use along with their perception of usefulness of the technology.

3.6. Healthcare quality perspective

The healthcare quality perspective addresses patient safety, appropriateness, and efficacy or health outcome. *Patient safety* concerns evaluations studying whether the use of the chatbot might create patient harm or risks for patients. For example, a medication assistant chatbot should provide the correct dosage of the medication to be taken. *Appropriateness* of using a healthcare chatbot assesses whether it is appropriate to deliver a certain healthcare service by means of a chatbot, e.g. whether it is appropriate to deliver cognitive behavior therapy using a chatbot for a particular patient. Finally, *health outcome or efficacy* has to be assessed aligned with practice of Evidence-Based Medicine (EBM). Then the ideal is 'Level 1' evidence as produced by randomized controlled trials (RCTs). To conduct an RCT requires measurement of a relevant validated health outcome as the dependent variable, and random assignment of subjects from the target population to the health chatbot or an appropriate 'control'. How to quantify the health outcome of a mobile application depends on the medical condition and treatment it is supposed to support (e.g. quantify efficacy of a diabetes management app by comparing the HbA1c value, or scoring systems such as Patient health questionnaire PHQ-9).

4. Discussion

This paper introduced an evaluation framework for health systems whose core interaction principle is a CUI. Jadeja et al. [10] distinguished four perspectives for evaluating general domain chatbots: Information retrieval (IR) perspective, UX perspective, linguistic perspective, and AI perspective. We adapted this categorization

by broadening the scope of the IR perspective to a task-oriented perspective. Furthermore, we included two dimensions that have not yet been reported explicitly for chatbot evaluation: the system quality perspective and the healthcare quality perspective. Low quality healthcare chatbots could readily harm their users in myriad ways - such as divulging confidential data, delaying available treatment, or recommending ineffective or directly contraindicated treatment - which must be avoided. Our framework suggests attributes to be assessed for a health chatbot. We still have to assess whether relevant evaluation aspects are included related to the Medical Devices Act that have to be considered for chatbots that makes recommendations for drug administration. What is still missing are experiences on the judgement, i.e. when can we consider a health chatbot to be good or appropriate to deliver health support for real patients? Depending on the scope of use of a chatbot, some of the quality attributes suggested in the framework might become irrelevant while others gain in importance. As a next step, we start a scoping review to fortify our framework. Afterwards, a Delphi study will be conducted to collect input on how to weight the different criteria. The ultimate goal of these efforts is to provide means that help in evaluating healthcare chatbots and to ensure that only validated, evidence-based, evaluated applications will be adopted by app stores or distributed by trustworthy distributors of health applications. Experiences with implementing the framework for chatbot evaluation has to be gained. For enabling comparison among healthcare chatbots with respect to quality, agreement upon standardized metrics for the single dimensions would be helpful.

References

- [1] L. Laranjo, A.G Dunn, H.L. Tong, et al., Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25(9) (2019), 1248-1258.
- [2] B. Inkster, S. Sarda, V. Subramanian, An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth* 6(11), e12106 (Nov 2018).
- [3] M. Tschanz, T.L. Dorner, J. Holm, K. Denecke, Using eMMA to manage medication. *IEEE Computer* 51(8) (2018), 18-25.
- [4] K. Denecke, S. Lutz Hochreutener, A. Pöpel, R. May, Self-anamnesis with a conversational user interface: Concept and usability study. *Methods Inf Med* 57(05/06) (2018), 243-252
- [5] K. Denecke, M. Tschanz, T.L. Dorner, R. May, Intelligent Conversational Agents in Healthcare: Hype or Hope? *Stud Health Technol Inform.* 259 (2019), 77-84.
- [6] H. Shum, X. He, D. Li, From Eliza to Xiaoice: Challenges and opportunities with social chatbots. *CoRR abs/1801.01957* (2018), <http://arxiv.org/abs/1801.01957>
- [7] E. Ruane, T. Faure, R. Smith, et al., Botest: A framework to test the quality of conversational agents using divergent input examples. In: *Proc. of the 23rd International Conference on Intelligent User Interfaces Companion*. IUI '18 Companion, ACM, New York, NY, USA (2018), 64:1-64:2.
- [8] R.B. Rubin, M.M. Martin, Development of a measure of interpersonal communication competence. *Communication Research Reports*, 11(1), (1994), 33-44.
- [9] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q* 13(3), (1989), 319-340.
- [10] M. Jadeja, N. Varia, Perspectives for evaluating conversational AI. *arXiv preprint arXiv:1709.04734*. 2017 Sep 14.

Implementation of an Open-Source Electronic Health Record for Decision-Support Education in Medical Informatics

Stephanie MEDLOCK^{a,1}, Roland SCHERMER^a and Ronald CORNET^a

^a*Amsterdam UMC, University of Amsterdam, department of Medical Informatics, Amsterdam Public Health Research Institute, Meibergdreef 9, Amsterdam, Netherlands*

Abstract. Access to electronic medical record systems is limited in many medical informatics education programs. The objective of this study was to inventory open-source patient record systems with decision support capabilities, implement a system for educational use, and test the effect of the system on students' learning. We sought systems that were under active development, with source code available, having an SQL-queryable database, and having decision support capabilities. We identified 20 candidate electronic health record systems, of which 6 mentioned decision support capabilities in their documentation. Of these, the OpenMRS system appeared to meet all of the requirements for use in our course; however, decision support capabilities needed to be added by use of a custom module implementing Arden2Bytecode, an Arden Syntax interpreter. Students who used this system showed an improvement in their knowledge of decision support systems and their capabilities. We conclude that there are a number of promising open-source electronic patient record systems currently under active development, but decision support capabilities are still immature. We anticipate further developments in this area in the coming years.

Keywords. Free and Open Source Software; electronic health record; decision support systems, clinical; medical informatics education

1. Introduction

Hospitals are moving from using a collection of small systems toward adoption of hospital-wide information systems. One of the major benefits expected from the use of electronic patient records is the ability to re-use data, for uses such as research and improving patient safety through clinical decision support systems (CDSS).

In an environment of many small medical software applications, Medical Informatics Master's students could readily build small tools which solved real information problems and gave them an active learning experience with systems similar to those that they will encounter in their career. Hands-on experience with an electronic patient record has been shown to be beneficial in health informatics education [1]. However, because most systems are closed-source, medical informatics students have no way to see how such systems are designed. This situation is further complicated by non-disclosure clauses, inhibiting the sharing of screen layouts and

¹ Corresponding Author. E-mail: s.k.medlock@amsterdamumc.nl.

interactions. Future medical informaticians will be expected to build decision support that is integrated with electronic health records; however, currently students can only get hands-on experience with stand-alone systems. With no patient record to use, students can only be educated in terms of knowledge and tested in terms of recall.

As a solution to the above problems, we propose creating a "virtual hospital" using one of the currently-available Free and Open Source electronic patient record systems. "Free and Open Source" (FOSS) means that not only is the software freely available, but the computer code which is used to make the software is also available, and is licensed so that anyone is free to make changes to the software. A number of such systems are available, including some that are in clinical use, particularly in middle-to-low-income countries [2]. However, it is not currently known if any such systems are suitable for use in medical informatics education, and particularly for teaching students about data re-use and building decision support applications.

The objective of this study was to choose a suitable FOSS patient record system with decision support capabilities, implement that system for use in Medical Informatics Master's education, and test the effect of the system on students' learning.

2. Methods

The first step was to select an electronic patient record for use. We searched for systems using an Internet search engine (duckduckgo.com) using synonyms for "open source" and "electronic patient record". This search was conducted on 1 June 2018. This was augmented by searching for lists of systems, using the names found in the first search. This process was repeated iteratively until no new systems were found. We then evaluated the documentation for these systems to assess for suitability for use in our educational program. The requirements were:

- The system should have source code available in a publicly-accessible repository.
- The repository should show current, active development (assessed by the number of contributors and commits in the last year).
- The system should be a hospital system, not a general practice system (systems which accommodate both would be accepted).
- The system should run on the Linux operating system.
- The system should have a database that is queryable in the Standard Query Language (SQL).
- Clinical decision support functionality should allow the creation of simple rule-based reminders.

The search and evaluation was carried out by one researcher (SM) with the results checked by a second researcher (RC).

Systems that appeared to meet the criteria were installed and assessed empirically to confirm that it met our needs. Once a system was selected, it was implemented on virtual machines and populated with synthetic patients. The data were generated to emulate the data quality problems found in actual patient record data: thus, the data set had missing data, data entry errors, misuse of coding systems, etc. Each group of students was made the "administrator" of a patient record system on a virtual machine. The 3-week assignment was to implement the recommendations for anticoagulant management from the ESC 2010 Guidelines for Atrial Fibrillation [3], assess the

accuracy of the advice generated by their system, and make a recommendation as to whether decision support should be implemented for the rest of the guideline.

Achievement of the learning goals was assessed through a voluntary pre- and post-test, which gave scenarios that CDSS implementers might encounter and asked the students to choose a course of action (Analyzing/Evaluating levels of Bloom's taxonomy). Students assigned to a different project served as non-randomized controls for students who completed this assignment. Statistical testing was not performed, due to the small number of students in the course.

3. Results

We identified 20 systems, of which 2 (OpenMRS and Bahmni) were judged to be suitable for use in our educational program. Bahmni is an extension of OpenMRS, thus we chose the core OpenMRS system for further investigation. Of the 20 systems identified in our search, 13 appeared to no longer be actively maintained (last commit > 1 year prior to the assessment). These systems were: FreeMed, OSCAR EMR, and openMAXIMS (last commit in 2017); ClearHealth, TolvenEMR, and UltimateEMR (last commit in 2016); Care2x (last commit 2015); WorldVisa and the Astronaut Vista installer (last commit 2014); OpenVista, LiUEEE (last commit in 2013); and OneTouchEMR and Solismed (no source code found). The other 7 systems, and the results of the assessment, are given in Table 1.

Table 1. Open-source electronic patient record systems identified by our search, and the results of assessment for suitability for decision-support education based on available documentation.

system	sourcecode repository	active development	hospital or GP	SQL variant	CDSS functionality
OpenMRS	github	303 contributors	hospital	MySQL	yes
Bahmni (based on OpenMRS)	github	78 contributors	hospital	MySQL	yes
OSEHRA Vista	github	3 contributors, 300 commits,	hospital	requires M-to-SQL	yes
OpenEMR	github	100 contributors	ambulatory	MySQL	yes
GnuHealth	gnu.org	2 contributors, 179 commits	hospital	PostgreSQL	maybe in an add-on module
OpenClinic GA	sourceforge	1 contributor, 22 commits	hospital	MySQL	unclear
Open Hospital	sourceforge	2 contributors, 80 commits	hospital	MySQL	unclear

All systems were cross-platform and thus met the criteria of being able to run on Linux virtual machines. Most systems were SQL-queryable, although derivatives of the Veteran's Administration's Vista system are written in the M programming language which incorporates a "no-SQL"-like database, and thus are not SQL queryable. Only a few systems had decision support functionality; the Vista-based systems, OpenEMR, OpenMRS, and Bahmni (which is based on OpenMRS). At the time, the decision support capabilities of OpenEMR were very simple, and it was designed for ambulatory rather than inpatient use.

Although the documentation for OpenMRS discussed support for Arden Syntax Medical Logic Modules (MLMs), an HL7-standard domain-specific language for

decision support, empirical evaluation showed no decision support capabilities in the current version. However, OpenMRS had good plugin support and used a standard interface (JDBC) to connect to its database. Therefore we were able to create a new OpenMRS addon module which supported the integration of Arden MLMs (see <https://github.com/RSSchermer/openmrs-module-ardenreminders>) using Arden2Bytecode, an existing FOSS Arden Syntax interpreter [4]. The system was hosted by a local service provider (Slash24 B.V.).

In the learning goals assessment, all 40 students completed the pre-test and 31 completed the post-test. Students who would later do the CDSS project (CDSS group, n=14) had a median pre-test score of 29% (IQR 15%-43%), while students doing other projects (non-CDSS, n=26) had a median pre-test score of 46% (IQR 34%-57%). On the post-test, the CDSS group showed a median absolute improvement of 10% (n=10, IQR 1-16%) while the non-CDSS group showed a median absolute improvement of 4% (n=21, IQR -4%-12%).

4. Discussion

We successfully incorporated an open-source patient record system with decision support in our Medical Informatics Master's program. The search for candidate systems identified 20 open-source patient record systems, of which 2 were suitable for use in our curriculum. Only 6 of the 20 systems had decision support capabilities; all of these were based on one of three core systems: OpenEMR, the Veteran's Administration VistA system, or OpenMRS. We were able to build a decision support module and implement it in OpenMRS. Assessment of students' mastery of the learning goals suggested that use of the system did aid in understanding the learning goals, although the number of students is too small to draw firm conclusions.

The strengths of our approach include a systematic assessment of the available candidate systems. However, the landscape of open-source systems is changing rapidly; this assessment should be considered a snapshot of the current state of the field. The main limitation is the small number of students in each group. Furthermore, in this year's class, we had 20 foreign exchange students, most with computer science backgrounds, participating in our class. By chance, a majority of the students in the CDSS group were foreign exchange students. This meant that only two regular Master's students completed both the pre- and post-test (with pre-test scores of 42% and 66%, and post-test absolute improvement of 12% and 15%). Although improvement for these two students was similar to that of the foreign exchange students, improvement in the CDSS group may still be attributable to the fact that the majority of these students had more room for improvement. However, it is encouraging that post-test scores on the learning goals assessment were acceptable even for students who started with little to no medical informatics background. We plan to continue to use the system and monitor its effect on learning goals, including the use of other methods for assessing learning goals.

A recent study assessed the capabilities of FOSS electronic health records, including their decision support capabilities [5]. The authors found 12 of the 20 systems that we found, plus an additional 5 systems not identified in this review (all of which appear not to be under current development or do not have source code available). They also found that OpenEMR and VistA were the systems that offered

decision support [5]. An earlier review of FOSS electronic health record systems did not assess decision support capabilities [6].

For our course, we used a custom-built decision support module integrated with OpenMRS. Previous efforts included the integration of an Arden Service, used by the Child Health Improvement through Computer Automation (CHICA) project [7]. Currently, OpenMRS does not have native decision support capabilities, but there are a number of promising efforts to add this functionality. For example, Kabukye et al. are working to develop a care pathway (including decision support) for oncology patients in low-resource settings [8]. Partners in Health is working to implement support for the triage of patients using the South African Triage System [9]. We intend to contribute to the development of other CDSS options, including integration of CDS Hooks and/or HL7 FHIR-based decision support, and we hope that future projects will include contributions from our own students to the OpenMRS code base

On the basis of this work, we conclude that there are a number of FOSS patient record systems under active development. However, the decision support capabilities of these systems are, at this time, limited. We successfully implemented an open-source patient record for teaching purposes: OpenMRS with a custom Arden Syntax module. Use of this system seemed to help students achieve the learning goals of the course. We look forward to future developments in decision support in open-source software.

The authors would like to thank the project funder, the Netherlands Initiative for Education Research Comenius program, project number 405.18865.274.

References

- [1] E.M. Borycki, J. Griffith, P. Reid, M.H. Kuo, A.W. Kushniruk. Do electronic health records help undergraduate health informatics students to develop health informatics competencies? *Stud Health Technol Inform.* **205** (2014), 838-42.
- [2] M. Alsaffar, P. Yellowlees, A. Odor, M. Hogarth. The State of Open Source Electronic Health Record Projects: A Software Anthropology Study. *JMIR.* **5** (2017):e6. doi:10.2196/medinform.578.
- [3] G.Y.H. Lip, J.L. Halperin, H-F Tse. The 2010 European Society of Cardiology Guidelines on the Management of Atrial Fibrillation. *CHEST.* **139** (2011), 738 - 741
- [4] M. Gietzelt, U. Goltz, D. Grunwald, M. Lochau, M. Marschollek, B. Song, K.H. Wolf. ARDEN2BYTECODE: a one-pass Arden Syntax compiler for service-oriented decision support systems based on the OSGi platform. *Comput Methods Programs Biomed.* **106** (2012), 114-25.
- [5] S. Purkayastha, R. Allam, P. Maity, J.W. Gichoya. Comparison of Open-Source Electronic Health Record Systems Based on Functional and User Performance Criteria. *Healthc Inform Res.* **2** (2019) 89-98.
- [6] A. Syzdykova, A. Malta, M. Zolfo, E. Diro, J.L. Oliveira. Open-Source Electronic Health Record Systems for Low-Resource Settings: Systematic Review. *JMIR.* **5** (2017) e44.
- [7] OpenMRS Talk pages. <https://talk.openmrs.org/t/how-can-i-call-an-arden-syntax-mlm-file-in-java/4697>. Visited 2019-9-5.
- [8] J. K. Kabukye, A.M. Casaceli, E. Ball, M. De Armas, R. Cornet. Towards an Open-Source Oncology Electronic Medical Records System for Low-Resource Settings: Development of Chemotherapy Management in OpenMRS. *Stud Health Technol Inform.* **264** (2019) 634-638.
- [9] Ball H. Point of Care in Haiti: Emergency Triage. <https://www.slideshare.net/lnball/openmrs-in-point-of-care-2017-p16-23>. Visited 2019-9-5.

Implementation of EXABO – An Expert Advisory Board for the European Reference Network for Rare Respiratory Diseases

Désirée WALTHER ^{a,1}, Olivia STEINMANN ^b, Thomas O.F. WAGNER ^b
and Holger STORF ^a

^aMedical Informatics Group (MIG), Frankfurt, University Hospital,
Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

^bFrankfurt Reference Center for Rare Diseases (FRZSE),
University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

Abstract. Rare conditions can make it difficult for physicians and healthcare providers to find the right answers and treatment for patients suffering rare diseases. The number of experts specializing in rare diseases in Europe is low. In order to ensure qualified diagnosis, treatment and care for these patients, the European Commission has founded the European Reference Networks (ERNs), which are virtual networks that include healthcare providers across Europe aiming to provide better care for patients across Europe. The European Reference Network LUNG (ERN-LUNG) has been established for patients with rare respiratory diseases seeking care and advice on all aspects of rare respiratory diseases. In the context of ERN-LUNG, the Expert Advisory Board (EXABO) was implemented. EXABO is a pan-European internet platform based on a question-answering system to support answering questions regarding all areas of rare respiratory diseases. This paper gives a brief overview of the first steps taken regarding the implementation of the EXABO platform, the challenges faced and the lessons that have been learned.

Keywords. Rare Diseases, rare respiratory diseases, ERN-LUNG, Expert Advisory Board

1. Introduction and Motivation

At present, the awareness of rare diseases is increasing. Diseases that affect less than 5 in 10,000 people are considered rare [1]. Between 6,000 and 8,000 rare diseases have been identified worldwide so far and affect 6% to 8% of the European population [1]. As a result, it becomes challenging for patients and their families to receive a confirmed diagnosis and adequate treatment for their condition, as the clinicians and healthcare providers often lack extensive experience in the majority of these rare conditions. Obtaining appropriate expert advice on a disease can be highly beneficial for patients as well as the clin-

¹Corresponding Author: Désirée Walther, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany.
E-mail: desiree.walther@kgu.de.

icians attempting to provide a targeted therapy. Launched in March 2017, the European Commission established the 24 European Reference Networks (ERNs) that work on a range of thematic issues [2]. Their main objective is to ensure and promote excellence in care and research for the benefit of patients affected by rare diseases. The European Reference Network for rare respiratory diseases (ERN-LUNG) addresses nine rare and complex pulmonary disease groups, including Interstitial Lung Disease (ILD), Cystic Fibrosis (CF), Primary Ciliary Dyskinesia (PCD), Pulmonary Hypertension (PH), nonCF-Bronchiectasis (nCF-BE), α 1-Antitrypsin Deficiency (AATD), Mesothelioma (MSTO), Chronic Lung Allograft Dysfunction (CLAD) and Other Rare Lung Diseases (ORLD), which are represented in nine Core Networks [3]. ERN-LUNG has 72 collaborating healthcare providers from 17 different European countries. Due to the lack of experts in the field of rare diseases, patients should benefit directly from expert knowledge, if they have a specific question regarding a rare respiratory disease. To provide the opportunity to address questions from patients and clinicians Europe-wide relating to any rare disease of the respiratory system to ERN-LUNG experts, the internet platform for an Expert Advisory Board (EXABO) was developed.

2. Methods

The concept of EXABO is based on the knowledge and experience with the former EU pilot project ECORN-CF², which is an online expert advisory board for Cystic Fibrosis. The key aspects of ECORN-CF were further developed and integrated into EXABO. While one area of rare respiratory diseases is covered in ECORN-CF, EXABO should be extended to all areas within ERN-LUNG. The experts must therefore be categorized based on their area of expertise and country. In each participating country, there is one expert group for each area of rare respiratory diseases. Each of these expert groups is organized by a moderator. The “superordinate moderator” takes on the task of organizing an area of rare diseases throughout Europe. In EXABO, the questions asked are distributed to the experts by moderators and the responses are monitored under quality aspects. Should an expert team not have the appropriate answer to a question, the superordinate moderator can reassign the question to the pool of Europe-wide experts [4].

Based on the stakeholder analysis and the concept, a prototype of EXABO was implemented. For the development of EXABO the waterfall model has been chosen. A two-staged testing process was used for testing. The results are significant for further development. In addition to the design and the technical realization of the Internet platform, the question-and-answer process, the internal communication between moderators and experts as well as the administration of the questions and users were of particular importance in the implementation of the prototype and the advancement to the present state. In the following the test phase is described and the improvements of the key features based on the evaluation of the test phase are highlighted.

2.1. Testing

The subject group, consisting of patients, relatives, physicians and programmers, was split up into the user roles questioner, moderator, superordinate moderator and administrator, so that a simultaneous interaction between them was possible.

²ecorn-cf.eu

In the quantitative test procedure, the subjects had to answer 30 questions regarding user friendliness, accessibility and navigation operating system based on a bipolar rating method using the four categories “do not agree”, “rather not agree”, “more likely agree” and “fully agree”. The neutral response option was excluded due to the fact that the test group for the EXABO prototype consisted of a very small number of persons.

The second stage was a qualitative test procedure, in which the subjects answered ten interview questions addressing improvement and weakness aspects.

2.2. Question-and-answer process

During the test phase, the question-and-answer process was tested for functionality and usability. The first step in the question-process in the prototype is selecting one of the eight rare and complex pulmonary conditions covered by ERN-LUNG: (1) ILD, (2) CF, (3) PCD, (4) PH, (5) nCF-BE, (6) AATD, (7) MSTO, (8) CLAD. In a next step, the questioner selects his desired language zone and inserts the question. The platform does not contain an automatic zone recognition, as it should be optional for the person using the EXABO platform to choose the language zone during travels. In the last step, personal data such as e-mail address, age, gender and country of residence are to be entered, whereby age and gender are not mandatory fields. For security reasons e-mail addresses are hidden even for logged-in users. The country of residence is essential for the assignment of questions as some languages are registered as official language in several countries. In consequence, a question that is asked in French will be sent either to a French or a Belgium group of experts, depending on the name of residency of the questioner.

The test phase has shown that a question cannot always be clearly assigned to a field of rare respiratory diseases in the beginning of the question-answering process. Above that, some questions may be so complex that it requires a different group of experts in the field of rare respiratory diseases. For this reason, the ninth Core Network “Other rare Lung Diseases” (ORLD) was set up. Furthermore, questions asked by a physician can be answered more specifically than those asked by laypersons, i.e. patients or their relatives.

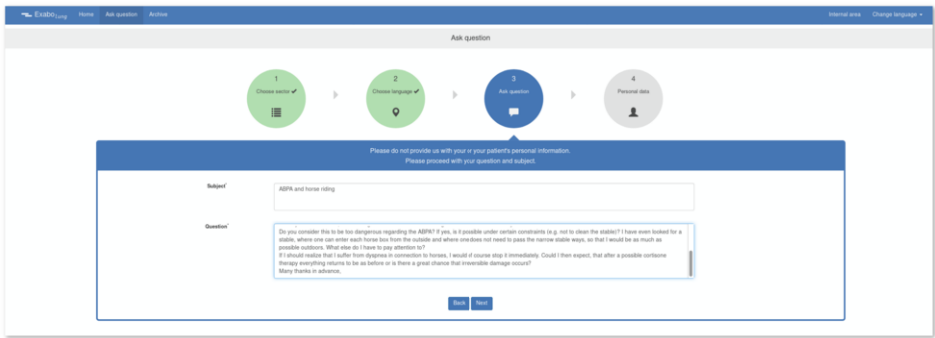


Figure 1. EXABO question entry mask.

After the questioning process, each question entered is automatically forwarded to the responsible moderator, who is the leader of his disease specific expert group and initiates the response process. The moderator assigns each incoming question to one of the ERN-LUNG experts in his group to either give the answer to the question or start

a discussion with other experts before publishing the response. However, questions can be returned by the expert if they are not fully answered. Returned questions are then translated to English and forwarded to the superordinate moderator, who will forward the question to the rare respiratory disease specific panel group across Europe, where the question can be discussed and the answer sent back to the respective moderator. When the answer is translated back in the source language, it can be published and archived on the internet platform.

During the test phase, it became apparent that a discussion with several experts is not only beneficial in the level of the superordinate moderator, but can also be initiated by the moderators in each expert group. Additionally, experts as well as moderators can comment on questions and answers and discuss the questions in a panel group of rare lung disease experts. In this context, the system provides date time for chat messages as well as a chat history, which enables the involved ERN-LUNG experts to retrieve their communication history. If the communication between an expert and the moderator involves additional experts and experts across Europe, the moderator may censor parts of the communication, i.e. if an expert has not fully answered a question and hence, returns it to the moderator. The reason for not having answered the question and other aspects can be set invisible by the moderator. However, the expert's commentary may already contain further approaches which are of advantage for other experts or discussion groups. In this case, the related moderator has the opportunity to censor the commentary instead of hiding the whole commentary.

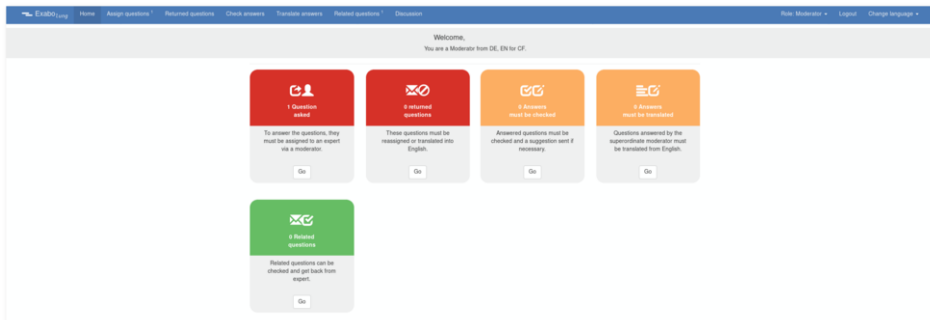


Figure 2. Moderator's welcome page with his jobs.

2.3. Administration of EXABO

An administrator manages the Internet platform and the user and role management. In the prototype implementation, the user interface of Keycloak [5] was used for the role and user management. The role distribution of EXABO is complex. In addition to experts, moderators and superordinate moderators, a user must also be assigned to a group of experts in a country with a specific field of expertise. Additionally, a user can be both an expert and a moderator, even for different fields of expertise or several countries. For example, a user can be an expert for cystic fibrosis and for primary ciliary dyskinesia, but is only the moderator for one of these fields. Accordingly, the responsible administrator who manages the user administration must work exactly to not forget or mix up a dependent role. If all roles are only listed without any structure, as is the case with the

prototype interface used, it will make it more difficult for the administrator to do his job correctly.

As an aid to the administrator, a completely new user interface with dependencies was programmed for the beta version. First, the actual role of expert, moderator, superordinate moderator and administrator is chosen, in the second step, the field of expertise is selected, whereby it is possible to make a multiple choice.

2.4. Technical realization

EXABO was implemented as a Spring Boot Java application and build with Maven. The front-end was realized with FreeMarker and Bootstrap [6]. The open-source single sign-on tool Keycloak [5] was used for the user and role management. The roles “experts”, “moderators” and “superordinate moderators” were included.

For the prototype a Web Application Resource (WAR) was used, so that an additional Tomcat server had to be installed on the server, in order to run the application. After the test phase and further development, it turned out that packaging with JAR and an integrated Tomcat server is the better choice for this application. The JAR is immediately executable.

3. Conclusion and Outlook

The generic aspect played a substantial role in the implementation of the EXABO Internet platform. As a result, new roles can easily be added in just a few steps and the platform can be enlarged for the implementation of other European Reference Networks.

The menu items are arranged in a logical manner, which makes the system user friendly and easy to use for laypersons as well as professionals. Translation into further European languages would allow a large number of other European citizens to enter their questions regarding rare lung diseases into the system and find more answers to their conditions. The internet platform is currently available in German and English in the beta version since March 2019 and is used by patients and care team members seeking advice on all aspects of rare respiratory diseases. The release of the full version is planned shortly. Special emphasis has been placed on a user-friendly interface/system/entry-points and an easy navigation operating system.

References

- [1] European Union Committee of Experts on Rare Diseases (EUCERD) Joint Action. State of the art of rare disease activities in Europe [Internet]. EUCERD Joint Action; 2014 [cited 2019 October 1]. Available from: <http://www.eucerd.eu/upload/file/Reports/2014ReportStateofArtRDAActivities.pdf>
- [2] European Commission. Conference Report. 3rd Conference on European Reference Networks; 2017 March 9-10 [conference proceedings on the Internet]. European Commission; 2017 [cited 2019 October 1]. Available from https://ec.europa.eu/health/sites/health/files/ern/docs/ev_20170309_frep_en.pdf
- [3] Humbert M, Wagner TOF. Rare respiratory diseases are ready for primetime: from Rare Disease Day to the European Reference Networks. *European Respiratory Journal*. 2017;49:1700085.
- [4] Walther D, Steinmann O, Schaefer J, Wagner TOF, Storf H. Conception of an Expert Advisory Board for the European Reference Network for Rare Respiratory Diseases. *Stud Health Technol Inform*. 2018;247:236-240.
- [5] redhat. (2017). *Keycloak*. Retrieved October 1, 2019, from <http://www.keycloak.org>
- [6] Otto, M., Thornton, J. (2017). *Bootstrap*. Retrieved October 1, 2019, from <https://getbootstrap.com>

Integrating a Secure and Generic Mobile App for Patient Reported Outcome Acquisition into an EHR Infrastructure Based on FHIR Resources

Nico PLONER^{a,1} and Hans-Ulrich PROKOSCH^a

^a *Department of Medical Informatics, Biometrics and Epidemiology, Chair of Medical Informatics, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Wetterkreuz 13, 91058 Erlangen, Germany*

Abstract. Both patients and physicians of the University Hospital of Erlangen would welcome means of digital data exchange between patients and the hospital. This work presents a system's architecture, FHIR based data model, security measures, and application flow based on patient's smartphones and a public cloud infrastructure to obtain patient reported outcome measures and merge that information with clinical routine data.

Keywords. Patient Reported Outcome Measures, FHIR, Mobile Applications

1. Introduction

Recent research revealed that both patients and physicians of the University Hospital Erlangen (UKER) would welcome means of digital data exchange between patients and the hospital [1]. Driven by the German eHealth law passed in 2015 [2] and encouraged by the growing research interest in patient reported outcome measures (PROM) [3-5], we developed the strategy for a digital patient connection into existing hospital IT infrastructures. For patients, the system relies on smartphone devices as a way of interacting with the hospital and their physicians. Information provided by the patient is pseudonymized, encrypted and stored in a public cloud infrastructure. From there it can be captured by a local application within the hospital's network, de-pseudonymized and then transmitted via FHIR interface to the hospital's electronic health record system (EHR).

Currently the system's purpose is to collect patient reported outcome measures, but it can easily be extended e.g. to allow patients to send other self-collected data (e.g. from wearable devices) to their physicians, or to enable physicians to send information to their patients while not being hospitalized. The hospital may specify definitions of care plans that describe at what point of time a patient is required to provide answers to a given questionnaire. Results are transmitted into the hospital's network and viewed by physicians.

¹ Corresponding Author, E-mail: nico.ploner@fau.de.

The objective of this paper is to describe the system’s generic data model, to illustrate the complete system architecture, including all security measures and to depict its use within a clinical scenario.

2. Methods

To enable a straightforward integration into existing systems, it was decided to use HL7’s Fast Healthcare Interoperability Resources (FHIR, [6]) as a data persistence and exchange format. First, the required subset of FHIR resources had to be identified and the corresponding data model had to be created. The smartphone application relies on this data model to its full extent. Participating hospitals may require a transformation into their proprietary data model when accessing patient data in the FHIR format provided by API interfaces.

Then, an architecture was created meeting the strong requirements of the German health care system regarding the exchange of patient-related data as well as granting access to any number of private smartphones without endangering the protected IT infrastructure of participating hospitals. This included the decisions on the location of the patient data storage, security measures to protect patient data against unauthorized access, mitigating the risk of leaked data, and providing access to participating hospitals in compliance with data protection regulations.

3. Results

3.1. The data model

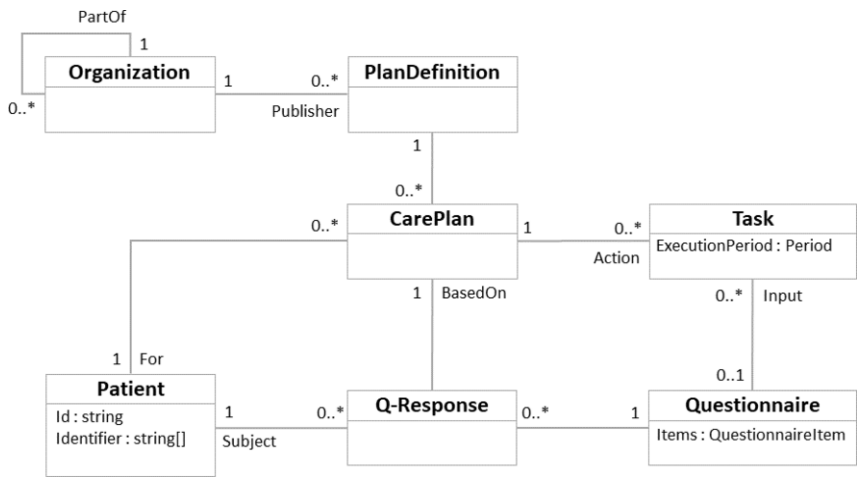


Figure 1. The system’s FHIR based data model

To facilitate interoperability, a subset of the standardized HL7 FHIR data model is used to store and exchange patient data. The seven FHIR resources utilized for this patient reported outcome application are linked together according to the data model illustrated in Figure 1.

Each participating department is represented by an "Organization" resource. Inner-hospital hierarchies can be mapped using the reflective relationship "PartOf". Each organization may publish one or more "PlanDefinition" resources which are used to define at what point of time a patient is requested to answer a questionnaire. Each patient is represented by a "Patient" resource with a unique Identifier ("Id", referred to as "Cloud ID"). A patient's pseudonym is stored in the "Patient" resource's "Identifier" collection. When an organization needs to collect data from a patient, a "CarePlan" resource is created that implements the specification of a particular "PlanDefinition" for that "Patient" and references all results. A "CarePlan" consists of multiple "Tasks", whereby a "Task" specifies at which point of time that patient needs to answer a given questionnaire. A "Questionnaire" resource consists of a list of questions to ask a patient and may serve as input to a "Task". A patient's answers to a given questionnaire are stored in a "QuestionnaireResponse" resource (abbreviated to "Q-Response" in Fig. 1) and subjected to the corresponding "Patient".

3.2. System Architecture

The entire system comprises of eleven components spread across four logical and three physical sites (see figure 2): The hospital's on-premise network infrastructure (A), a public cloud (B and C) and the patient's private smartphone (D).

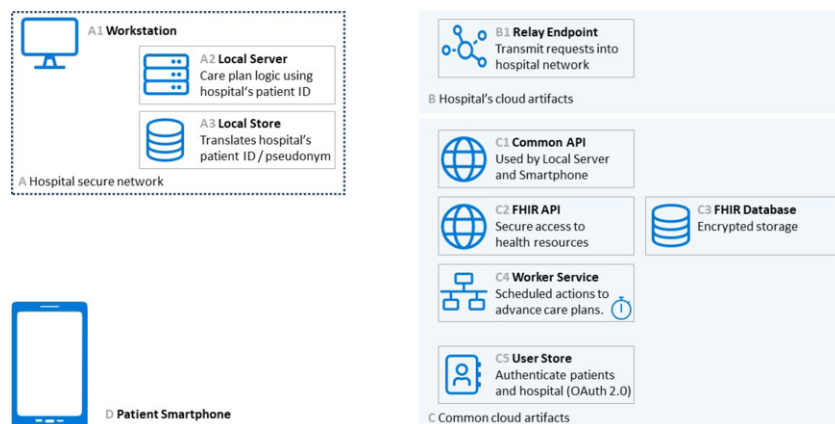


Figure 2. Integrating the patient app, the hospital's EHR and public cloud storage

Inside the hospital's secure network infrastructure (A) the system relies on the established workstations (A1) as user interfaces. Authenticated physicians may interact with the system to enroll patients into care plans and view patient reported results. Workstations connect to the system using a web server appliance (A2) hosted inside the hospital's network providing the patient's ID of the Hospital Information System. A local store (A3) is used for generating and managing unique pseudonyms for each patient. When communicating with the public cloud components (B and C) the localserver only uses that pseudonym, the patient's ID of the Hospital Information System is never transmitted outside the hospital.

No public endpoint into the hospital's network is exposed. The Local Server regularly polls the Relay Endpoint (B1) inside the public cloud for patients' responses to questionnaires.

In order to provide access to both, the hospital and its patients, a public cloud infrastructure is used. That infrastructure is located on German soil and run by a German provider subjected to the General Data Protection Regulation (GDPR)[7].

Common artifacts are services used by both, the Local Server (A2) hosted inside the hospital's network infrastructure and the patient's smartphone (D).

A common API (C1) provides functionality like creating a new patient record by the smartphone application (D), or requesting a one-time code for patient enrollment to be presented at the hospital's workstation by the Local Server (A2). Patient data is stored inside a FHIR database (C3) on an encrypted storage. A FHIR API (C2) provides secure access to the patients' data. A scheduled Worker Service (C4) regularly checks for updates in care plans and requests patients to answer questionnaires or transmits responses to the Relay Endpoint (B1). A cloud User Store (C5) and authentication service is used to secure communication between components and limit access to patients' data. Patients may interact with the system using their private smartphone device. A platform specific application is available in the usual app stores. The use of the respecting application does not require the patient to disclose identifying features such as name, email, or telephone number. It is also refrained from using device specific identifiers like IMEI, SSAID, MAC addresses/EUI, or Advertising-IDs. Instead, a random installation ID is created and used to create a new user in the User Store (C5) and a new patient resource in the FHIR Database (C3) after downloading the app.

3.3. Security

The following measures are taken to protect data and secure communication with and within the system:

- i. Data minimization: Patients using the smartphone application are not forced to disclose identifying information.
- ii. Pseudonymization: Information exchange between the components hosted inside the hospital's network relies on randomized pseudonyms preventing local identifier and cloud ID from being disclosed.
- iii. Short-term access codes: In order to enroll in a care plan, the patient enters a code displayed at the physician's workstation into the smartphone application. This code is valid for only a single enrollment within a short period of time.
- iv. Secure communication: For all communication between components, Transport Layer Security (TLS/SSL) is enforced.
- v. OAuth 2.0 for access: Access to a patient's data is restricted to their smartphone and to confidential clients (i.e. cloud components).
- vi. Encryption-at-rest: When saved to the FHIR Database, all data is encrypted so that even with theft of the data media no access is possible. All other cloud components hold decrypted data only in transient memory and never persist in health information.
- vii. Relay into hospital network: To deliver information into the on-premise hospital infrastructure, no public endpoint into the hospital's network is exposed. Using a polling mechanism, the local server reads data from the relay endpoint containing information to be transferred into the hospital.

3.4. Exemplary usage

The patient downloads and installs the application on his smartphone. After download, a random installation ID is created on that patient's smartphone. In the public cloud, a new "Patient" resource with a new Cloud ID is created and stored in the FHIR Database that may only be accessed by that patient's smartphone. The physician opens

the patient's health record in the EHR system and requests enrollment to a care plan. The EHR then requests a one-time code for enrollment from the Local Server providing the patient's ID. The Local Server either retrieves that patient's pseudonym from the Local Store or creates a new one. The Local Server then requests the one-time code from the Common API providing the pseudonym and an identifier for the care plan and passes the one-time code to the workstation for display.

The patient enters that one-time code into the smartphone application and chooses to enroll in this care plan. The smartphone application requests enrollment at the Common API, which creates the necessary resources in the FHIR database. Invoked by a schedule, the Worker Service determines, that this patient is required to answer a questionnaire and sends a push notification to the patient's smartphone. The patient then may provide answers in a given period of time. The results are stored in the FHIR Database. Invoked by a schedule, the Worker Service determines that the patient provided answers to a questionnaire. Using the Patient's pseudonym, the Worker Service posts a corresponding message to the hospital's Relay Endpoint. The Local Server polls the message from the Relay Endpoint, resolves the patient's ID in the Hospital Information System, stores the patient's answers and may notify the physician.

4. Conclusions

In this paper, we described a system relying on patients' smartphones and a public cloud infrastructure to create a digital connection into existing hospital IT infrastructures. An assessment by the data protection officer and the IT security officer at the University Hospital Erlangen already granted permission to use the system in studies using real patient data. We are currently planning initial research projects with departments at the University Hospital Erlangen in which the system is to be used and evaluated.

References

- [1] N. Ploner, et al., "Concept to gain trust for a German Personal Health Record system using Public Cloud and FHIR", *Journal of biomedical informatics* (2019), 103212.
- [2] German law: "Gesetz zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz - GMG) vom 14. November 2003", 2003, BGBl. I S. 2190.
- [3] R.E. Jensen, et al., "The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care." *Medical care* **53**(2) (2015), 153.
- [4] T.W. LeBlanc, et al., "Patient-reported outcomes in cancer care - hearing the patient voice at greater volume", *Nature reviews Clinical oncology* **12** (2017), 763.
- [5] P. Leiner, "Onkologen setzen auf die Eigenwahrnehmung", *ÄrzteZeitung online*, 15.12.2017 <https://www.aerztezeitung.de/medizin/krankheiten/krebs/article/954349/patienten-kennen-besten-onkologen-setzen-eigenwahrnehmung.html> (Accessed September 28, 2019).
- [6] M. Lehne, et al., The Use of FHIR in Digital Health - A Review of the Scientific Literature. *Stud Health Technol Inform.* (2019), Sep 3;267:52-58.
- [7] EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679, OJ 2016 L119/1

Interaction with Medical Venues in Megacities

María ÓSKARSDÓTTIR ^{a,1} and Anna Sigríður ÍSLIND ^a

^a*Department of Computer Science, Reykjavik University, Iceland*

Abstract. With the vast amounts of data that is being generated today, come new possibilities of understanding patient mobility. In this study, of urban mobility in ten mega cities worldwide, we try to understand the relationship between patients' environment and behaviour with regards to venues that provide some kind of medical care. We analyze longitudinal mobility data set from ten of the world's megacities and investigate urban dynamics and travel patterns in terms of interaction with medical centers. Our goal is to investigate universal patterns and gain an understanding of where people are coming from when they visit such venues and where they go afterwards as well how travel patterns progress throughout the day.

Keywords. Mobility, Health Data Science, Location Based Social Networks, Patient Behaviour

1. Introduction

Due to the global environmental changes and increased focus on sustainable living, there is a growing interest in understanding the movement patterns of people [1]. In line with that, the research on travel medicine and epidemiology has benefited from the increase of technologies such as location based social networks (LBSN) that enable geotagging as a new way of understanding mobility patterns and urban behaviour. The research on LBSN has been focused on understanding different aspects of mobility and user behavior [2,3,4]. Within that interest is mobility and user behaviour in relation to health purposes. Research has e.g. focused on geographic variation in well-being in Tweets [5] and exploring Twitter for medicine and health-related terms to find geographical patterns in symptoms control [6]. Furthermore, Foursquare data has been analysed to gain an understanding of the spatial distribution of ambulance incidents [7]. There is however still a gap in the literature regarding the understanding of how people interact with medical venues. In this paper we explore what people do before and after they visit a medical center, when their visits take place, and whether the mobility patterns are universal.

We analyze a longitudinal mobility data set from the location based social network service Foursquare, which has over 50 million active users monthly. We explore the movements of people between venues in the cities Chicago (CI), Istanbul (IS), Jakarta (JA), London (LO), Los Angeles (LA), New York (NY), Paris (PA), Seoul (SE), Singapore (SI) and Tokyo (TO). The mobility data set originates from Foursquare and was

¹Corresponding Author: María Óskarsdóttir, School of Computer Science, Reykjavik University, Menntavegi 1, 101 Reykjavik, Iceland; E-mail: mariaoskars@ru.is.

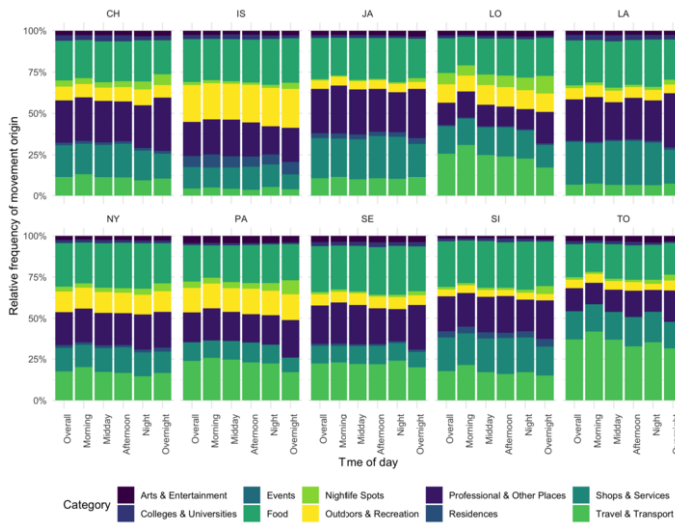


Figure 1. Relative frequency of categories of the origin of movements when destination is Medical Center

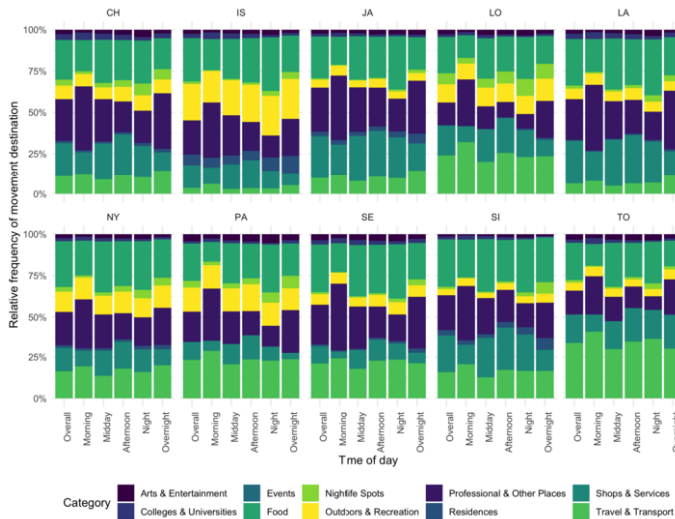


Figure 2. Relative frequency of categories of the destination of movements when origin is Medical Center

provided for the participants of the Future Cities data challenge² and spans two years. A movement is created in the data set each time a user checks-in at one of the city's venues and consists of two end points, the origin, where the user was last time they checked in and the destination where the user is currently checking in, as well as the part-of-day indicating the time of arrival at the destination, defined as morning (06-10), midday (10-15), afternoon (15-19), night (19-24) and overnight (00-06).

²See <https://www.futurecitieschallenge.com/>

Foursquare uses a hierarchical category set to categorise the venues on the platform. There are ten categories, as shown on the legend in Figure 1. Each category is separated into subcategories which are furthermore divided into subsubcategories. Medical Centers are for instance a subcategory within the category Professional & Other Places, and can be separated into subsubcategories such as Hospitals, Dentist's Offices and Physical Therapists. In total there are 923 categories³ on different levels where we, in this paper, focus on venues that belong to the subcategory Medical Centers.

2. Analysis of Movements

For our analyses, we inspect the data from two perspectives, namely where people are coming from when they visit medical venues and where they go to after their visits, by extracting all movements that end and start in a venue in the Medical Center subcategory, respectively. We look at the distribution of origin and destination venues on two levels of granularity, that is, category and subcategory.

Figure 1 shows the distribution of categories of the movement origins in the ten cities. In each subplot, there are six bars showing the distribution during various parts of the day. The first bar in each subplot shows the overall distribution, that is, irrespective of the time of day, where as the remaining five columns show, from left to right, the distribution in the morning, at midday, in the afternoon, at night and overnight. Evidently, this distribution varies from city to city, for example, venues in the category Outdoors & Recreation are more frequently visited in Istanbul than in Jakarta, and Travel & Transport venues are used more in Tokyo than in Chicago and Los Angeles. There is also a difference in patterns throughout the day. In London, for example, Travel & Transport venues have a higher frequency in the morning and in many cities the frequency of Nightlife Spots is the highest in the overnight period. The variation in venue categories is more prominent in Figure 2 which shows the distribution of to movements. Here, we can see that in many of the cities, Professional & Other Places are much more often travelled to from Medical Centers in the morning than overall, Food venues are less frequented in the morning and more at night.

Next we look at the venues travelled from and to at the more granular subcategory level. Based on Figures 1 and 2 we select the cities New York and Tokyo, because they show variety in their distribution of movements and they represent cities in different parts of the world. As there are over four hundred subcategories, we only look at the most frequented ones. Tables 1 and 3 show the highest ranked subcategories of venues of movement origins in New York and Tokyo at different parts of the day, respectively. Tables 2 and 4 show the same for the destination of movements. The first entry in each column is the most frequent subcategory during the respective part-of-day, and so on. The numbers in brackets show the ranking of the subcategory in the overall data set, that is, taking into account all movements. This gives an indication of the general frequency of the subcategories in the data set. These tables give some interesting insights into how people interact with Medical Centers within and between the two cities. First of all, Medical Centers appear very high in all the tables. This means that movements where both end points belong to Medical Centers are very common. This is probably due to people who are employed at these venues and therefore go there regularly and frequently check-in.

³See <https://developer.foursquare.com/docs/resources/categories> for a complete list

Table 1. The highest ranked subcategories of venues visited before a Medical Center in New York.

Morning	Midday	Afternoon	Night	Overnight
1 Medical Centers	Medical Centers	Medical Centers	Medical Centers	Medical Centers
2 Metro Stations (1)	Metro Stations (1)	Metro Stations (1)	Metro Stations (1)	Metro Stations (1)
3 Food & Drink Shops (3)	Parks (2)	Food & Drink Shops (4)	Food & Drink Shops (5)	Parks (6)
4 Parks (4)	Food & Drink Shops (4)	Parks (2)	Parks (4)	Bars (1)
5 Train Stations (2)	Coffee Shops (3)	Coffee Shops (3)	Bars(2)	Food & Drink Shops (7)
6 Coffee Shops (6)	Athletics & Sports (5)	Athletics & Sports (6)	Athletics & Sports (6)	Athletics & Sports (9)
7 Athletics & Sports (5)	Pharmacies	Asian Restaurants (8)	Asian Restaurants (3)	Asian Restaurants(5)
8 Pharmacies	Asian Restaurants (7)	Bars (5)	Coffee Shops (7)	Train Stations (8)
9 Bars (8)	Train Stations (8)	Pharmacies	Pharmacies	Pharmacies
10 Asian Restaurants	Bars (6)	Train Stations (7)	Fast Food Restaurants	Residential Buildings

Table 2. The highest ranked subcategories of venues visited after a Medical Center in New York.

Morning	Midday	Afternoon	Night	Overnight
1 Medical Centers	Medical Centers	Metro Stations (1)	Medical Centers	Medical Centers
2 Metro Stations (1)	Metro Stations(1)	Medical Centers	Metro Stations(1)	Metro Stations(1)
3 Parks (6)	Food & Drink Shops (4)	Food & Drink Shops (2)	Food & Drink Shops (4)	Parks (5)
4 Coffee Shops (2)	Parks (3)	Parks (3)	Bars (2)	Bars (2)
5 Athletics & Sports (3)	Coffee Shops (2)	Athletics & Sports (7)	Asian Restaurants (3)	Athletics & Sports (6)
6 Food & Drink Shops (8)	Athletics & Sports (5)	Coffee Shops(6)	Parks (5)	Airports (3)
7 Train Stations (4)	Asian Restaurants (6)	Pharmacies	Athletics & Sports (7)	Bridges
8 Offices (5)	Pharmacies	Bars (4)	Pharmacies	Fast Food Restaurants
9 Donut Shops	Offices (7)	Train Stations (5)	Train Stations (8)	Food & Drink Shops (9)
10 Bridges	Fast Food Restaurants	Asian Restaurants (8)	Coffee Shops (10)	Train Stations (7)

Table 3. The highest ranked subcategories of venues visited before a Medical Center in Tokyo.

Morning	Midday	Afternoon	Night	Overnight
1 Train Stations (1)	Train Stations (1)	Train Stations (1)	Train Stations (1)	Train Stations (1)
2 Asian Restaurants (3)	Asian Restaurants (2)	Asian Restaurants (2)	Asian Restaurants (2)	Medical Centers
3 Metro Stations(2)	Metro Stations (3)	Medical Centers	Metro Stations (3)	Asian Restaurants (2)
4 Medical Centers	Medical Centers	Metro Stations (3)	Medical Centers	Convenience Stores (5)
5 Convenience Stores (4)	Convenience Stores(4)	Convenience Stores (4)	Convenience Stores (5)	Metro Stations (3)
6 Food & Drink Shops (5)	Food & Drink Shops (6)	Food & Drink Shops (7)	Food & Drink Shops (8)	Bars (4)
7 Coffee Shops (6)	Coffee Shops (5)	Coffee Shops (6)	Coffee Shops (6)	Parks
8 Intersections (8)	Parks	Shopping Malls (5)	Intersections	Food & Drink Shops (6)
9 Parks	Intersections (10)	Parks	Bars (4)	Intersections (7)
10 Bridges (7)	Shopping Malls (7)	Intersections	Shopping Malls (7)	Coffee Shops (8)

Table 4. The highest ranked subcategories of venues visited after a Medical Center in Tokyo.

Morning	Midday	Afternoon	Night	Overnight
1 Train Stations (1)	Train Stations (1)	Train Stations (1)	Train Stations (1)	Train Stations (1)
2 Medical Centers	Asian Restaurants (2)	Asian Restaurants (2)	Asian Restaurants (2)	Medical Centers
3 Metro Stations (2)	Medical Centers	Metro Stations(3)	Metro Stations (3)	Convenience Stores (3)
4 Coffee Shops (5)	Metro Stations (3)	Medical Centers	Convenience Stores (5)	Asian Restaurants (2)
5 Convenience Stores (3)	Convenience Stores (4)	Food & Drink Shops (4)	Food & Drink Shops(6)	Metro Stations (4)
6 Asian Restaurants (4)	Coffee Shops (5)	Convenience Stores (6)	Medical Centers	Food & Drink Shops (8)
7 Schools	Food & Drink Shops (7)	Coffee Shops (7)	Bars (4)	Parks (10)
8 Intersections (8)	Intersections (10)	Shopping Malls(5)	Intersections(9)	Intersections (7)
9 Parks	Shopping Malls (6)	Intersections	Coffee Shops (8)	Hotels (6)
10 Bridges (6)	Parks (9)	Cafés (9)	Shopping Malls (7)	Steakhouses

Secondly, Train and Metro Stations also appear very frequently. We already know that Travel & Transport venues are common in these two cities, and as with other venues, people use public transport when commuting to and from Medical Centers. Places for eating and buying food are also dominant in combination with Medical Centers. When comparing the two cities, we can highlight a few aspects. In New York, Pharmacies have a high ranking both before and after visiting a Medical Center. This is not the case in the other cities. We also see that venues that offer food and are not ranked high in general, appear in the list after visiting a Medical Center in New York, such as Fast Food

Restaurants and Donut Shops. In Tokyo, there is high correspondence between general travel patterns and those related to Medical Centers, as seen by the numbers in the brackets. Only Parks seem to be out of place here, as they are highly ranked before visiting a Medical Center.

3. Conclusion

Data from LBSN have the potential of revealing different patterns in human mobility. In this paper we draw from a study where we explore mobility patterns on what people do before and after their visit to a medical center, based on data from the LBSN Foursquare. We compared data from ten megacities and examined two megacities in more detail.

As the data set is from ten megacities, it should make the users rather homogeneous in the different cities. However, the different cities exhibit heterogeneous travel patterns where different categories of venues are dominant for each city. For instance, the sub-category of athletics and sports is absent in the Tokyo data but is frequent in the New York data. Another interesting finding is the lack of pharmacies. The New York data shows that pharmacies are visited both before and after visits to a medical center, which is not the case in the other cities. The different user behaviour can also be shown through apparent interaction with bridges and intersections in the Tokyo data set in general and with parks in relation to a visit to a medical center in particular. As the data comes from the same period of time, Foursquare as a LBSN looked and performed the same at all locations. However, the popularity in the different places and the stream of tourists that either use, or do not use LBSN, can vary in the ten megacities. Future work could both include understanding what the data represents in terms of the users of the LBSN, and to understand the patient patterns. Furthermore we want to examine the distance between the venues that show up in the data and explore the different ways geotagging is used in the ten megacities in relation to visits to medical centers and develop a measurement of the similarities between the geotagging of medical centers and overall categories.

References

- [1] Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *nature*, 453(7196), 779.
- [2] Deeva, G., De Smedt, J., De Weerd, J., & Óskarsdóttir, M. (2019, December). Mining Behavioural Patterns in Urban Mobility Sequences Using Foursquare Check-in Data from Tokyo. In *International Conference on Complex Networks and Their Applications* (pp. 931-943). Springer, Cham.
- [3] Silva, T. H., Viana, A. C., Benevenuto, F., Villas, L., Salles, J., Loureiro, A., & Quercia, D. (2019). Urban computing leveraging location-based social network data: a survey. *ACM Computing Surveys (CSUR)*, 52(1), 17.
- [4] Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: universal patterns in human urban mobility. *PloS one*, 7(5), e37027.
- [5] Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Lucas, R. E., Agrawal, M., ... & Ungar, L. (2013, June). Characterizing geographic variation in well-being using tweets. In *Seventh International AAAI Conference on Weblogs and Social Media*.
- [6] Paul, M. J., & Dredze, M. (2011, July). You are what you tweet: Analyzing twitter for public health. In *Fifth International AAAI Conference on Weblogs and Social Media*.
- [7] Noulas, A., Moffatt, C., Hristova, D., & Gonçalves, B. (2018, April). Foursquare to the rescue: Predicting ambulance calls across geographies. In *Proceedings of the 2018 International Conference on Digital Health* (pp. 100-109). ACM.

Investigating the Barriers to Physician Adoption of an Artificial Intelligence-Based Decision Support System in Emergency Care: An Interpretative Qualitative Study

Cécile PETITGAND,^{a,b,1} Aude MOTULSKY^{a,c}, Jean-Louis DENIS^{a,c}, and Catherine RÉGIS^{b,a}

^aResearch Center, Centre hospitalier de l'Université de Montréal, Montreal, Canada

^bFaculty of Law, Université de Montréal, Montreal, Canada

^cDepartment of Management, Evaluation and Health Policy, Université de Montréal, Montreal, Canada

Abstract. The development of artificial intelligence (AI) systems to support diagnostic decision-making is rapidly expanding in health care. However, important challenges remain in executing algorithmic systems at the frontlines of clinical practice. Hence, most often, these systems have not been trained with local data nor do they fit with context-specific patterns of care. This research examines the implementation of an AI-based decision support system (DSS) in the emergency department of a large Academic Health Center (AHC) in Canada, focusing specifically on the question of end-user adoption. Based in an interpretative perspective, the study analyzes the perceptions of healthcare managers, AI developers, physicians and nurses on the DSS, so as to make sense of the main barriers to its adoption by emergency physicians. The study points to the importance of considering interconnections between technical, human and organizational factors to better grasp the unique challenges raised by AI systems in health care. It further emphasizes the need to investigate actors' perceptions of AI in order to develop strategies to adequately test and adapt AI systems, and ensure that they meet the needs of health professionals and patients. This research is particularly relevant at a time when considerable investments are being made to develop and deploy AI-based systems in health care. Empirically probing the conditions under which AI-based systems can effectively be integrated into processes and workflow is essential for maximizing the benefits these investments can bring to the organization and delivery of care.

Keywords. artificial intelligence, decision support system, emergency care, adoption

1. Introduction

The development of artificial intelligence (AI)-based systems to support clinical decision-making in health care is currently undergoing rapid expansion. AI, broadly

¹ Corresponding Author, Cécile Petitgand, Email: cecile.petitgand@umontreal.ca.

defined as the imitation of human cognition by a machine [1], is expected to enable better surveillance, detection and diagnosis of illnesses, as well as uncover novel treatments to spur precision medicine [2-4]. AI systems have already proven more effective than dermatologists at diagnosing skin cancer [5], and experts believe they will outperform medical specialists in surgery by 2050 [6]. As a result, health care is often presented as one of the most propitious domains in the upcoming "AI revolution" [7][8]. Yet, despite the enthusiasm around AI's capacity to increase the quality, safety and efficiency of care, important challenges remain in executing algorithmic systems at the frontlines of clinical practice. Hence, most often, these systems have not been trained with local data, nor do they fit with context-specific patterns of care [1][9][10].

The present research aims to explore some of these challenges by analyzing the implementation process of an AI-based decision support system (DSS) in the emergency department of a large Academic Health Center (AHC) in Canada. Research to date has identified several hurdles facing the deployment of AI systems in health care. These relate mainly to technical issues (e.g. lack of quality data sets to train algorithms; non-interoperable information platforms) and ethical/legal considerations (e.g. concerns about data privacy, algorithm opacity and patient safety; lack of ethical and legal frameworks to provide safeguards against inappropriate use) [11-14]. Research gaps remain in understanding how these barriers interrelate with human and organizational factors to drive and influence implementation.

This article adopts an interpretative perspective [15] to analyze the implementation of an AI-based DSS in an emergency department (ED), focusing on actors' representations of the system. We explore how health managers, AI developers, physicians and nurses perceive the AI-based DSS and interpret the interrelated technical, human and organizational barriers that lead to its ineffective adoption in the AHC.

2. Methodology

2.1. Study site and AI system

The study site is one of the largest AHCs in Canada, located in the province of Québec. In November 2018, an AI-based DSS was tested for implementation as a pilot project in the AHC's ED. The DSS is a diagnostic technology based on a deep-learning algorithm using clinical data and evidence-based results from the scientific literature. For patients visiting the ED, the DSS presents as a questionnaire that is answered on a mobile tablet. First, the patient is asked about the purpose of their visit to the ED. A Natural Language Processing engine then analyzes the text written by the patient to identify the chief complaint (e.g. skin rash, abdominal pain, etc.).

The patient is then asked a series of questions based on the chief complaint, with each question adapted to the previous response. The main objective is to identify red flags and signs of serious conditions. At the end, the DSS outputs a medical history that is printed by nurses or clerks and presented to the physician prior to their encounter with the patient. According to the DSS developers, who are also emergency physicians at the AHC, the medical history is meant to optimize patient questioning and diagnostic decision-making by providing physicians with information about the history of the presenting complaint, pertinent positives and negatives, past medical/surgical/family history, and social history.

2.2. Data collection and analysis strategies

An in-depth case study was performed at the ED, triangulating data collection methods and sources in an ethnographic approach to make sense of actors' perceptions of the DSS implementation process. A snowball technique was used to recruit participants and 20 semi-structured interviews were conducted: first with the DSS developers (5) and AHC managers (5), and then with emergency physicians (7) and nurses (3). Complementing the interviews (which were recorded and transcribed), informal conversations with nurses and clerks took place, as well as non-participant observations of meetings with managers, designers, physicians and nurses (10 in total), and of DSS utilization in context (15 hours). Field notes allowed observed situations to be contextualized, and preliminary elements of analysis to be identified. Finally, multiple secondary documents were gathered to better grasp actors' interpretations and trace the DSS implementation history; these included emails sent to physicians, documents prepared by AHC managers and developers, as well as different versions of the DSS-generated medical history. For data analysis, a description of the implementation process was first completed using a narrative strategy [16] that highlighted key timelines and activities. Then, using all collected data, thematic content analysis was performed with NVivo 12 software (QSR International) to identify, categorize and refine the main barriers to physician adoption of the DSS. [17]

3. Findings

Our analysis of the implementation process revealed several barriers identified by research participants that limited the DSS integration into clinical processes and workflow, and eventually drove several physicians towards non-adoption. These barriers related to three types of issue.

3.1. Availability

First, trials were conducted with patients to test whether they could complete the DSS questionnaire. Following adjustments to improve the intelligibility of the questionnaire, AHC managers and DSS developers jointly decided that only a subsection of patients would be asked to complete the questionnaire: English- or French-speaking outpatients who did not present any mental deficiency, visual handicap or alcohol/drug intoxication symptoms. This considerably limited the number of medical histories generated by the DSS.

Moreover, the lack of interoperability between the DSS and the AHC clinical information systems (Electronic Patient Record and Emergency Information System) meant that medical histories had to be printed and handed to physicians in paper form. These tasks were assigned to nurses and clerks who were already overworked with their regular obligations. As a result, medical histories were often not printed and thus were not provided to physicians. According to the DSS developers, this was the main factor limiting physician adoption of the medical history. Developers argued that physicians did not have enough opportunities to read medical histories and learn how to integrate them into their clinical practice. However, according to nurses, this did not really explain why physicians were not becoming "early AI adopters":

Every time it doesn't work, they say it's because nurses didn't print the history, or didn't direct them [patients] to the tablet, or because... "Wait a minute, you, [developers], what have you done to change our processes? What have you done to improve the medical history? What have you done to innovate?"[...] Since, it's something to help physicians, they need to play their part. (Nurse 2)

3.2. Usability

Some nurses were convinced that physicians were not using the medical histories because they had difficulty understanding patient information reported by the DSS. This was, in fact, a major barrier reported by several physicians. The physicians interviewed considered that the AI-based system was good at reporting simple complaints (a localized pain, a broken leg, etc.) but very poor at making sense of multi-complaint conditions (pain throughout the body, pain related to severe pre-existing conditions, etc.). This was a major concern, as most patients coming to the ED presented with the latter profile:

The history shows a multitude of symptoms to which the patient responded 'Yes, I have this'. But is it relevant? Is it active? Is it related to the current complaint? Afterwards, you need to disentangle all this. (Physician 4)

This type of feedback was directly reported to the DSS developers during focus groups. In response, adjustments were made to the design of the medical history to better classify patient information and simulate the clinical reasoning of physicians (presentation of pre-existing conditions, chief complaint, etc.). However, adoption rates did not increase significantly. Implementation data collected by the DSS developers showed that the proportion of annotated medical histories remained almost the same (around 30%) before and after these adjustments.

3.3. Perceived usefulness

At the start of implementation, several physicians were positively disposed toward using an AI-based DSS to enhance their diagnostic practice. However, some reported having discovered "errors" in the medical histories. In particular, two physicians reported that reading the medical history led them down the wrong diagnostic path. Had they not questioned the patient again, they would have made a serious clinical error:

But, you know, there are times when it completely took me down the wrong path... Not often... But it happened and it should not happen. It's like, me... I have zero tolerance. It's a tool that's supposed to help us... not at the price of losing a patient... But to miss something, something huge, you know... So that's... It didn't happen often. But it happened, so it cooled my enthusiasm. (Physician 6)

Perceptions of DSS-induced errors were shared among physicians, and this led some to develop a persistently sceptical attitude towards the usefulness of the DSS. The AI system was thus viewed as introducing a real risk into clinical practice that was capable of causing harm to patients. This is a perception that tends to increase clinician resistance to health information systems.

4. Discussion and implications

This in-depth case study shows how the combination of availability, usability and perceived usefulness can contribute to physician distrust of an AI-based DSS. The study further points to the importance of empirically probing the interconnections between technical, human and organizational factors to make sense of barriers that limit the implementation of AI-based systems in health care. However, these interconnections are rarely taken into account by researchers working on the deployment of AI-based technologies in clinical environments [4][9].

Moreover, the research emphasizes the need to consider the unique challenges raised by AI integration into clinical processes and workflow [10]. As shown in this study, actors' perceptions of a technology influence their actions towards it, be they related to adoption or resistance. For this reason, it is essential that managers responsible for implementation deal with specific assumptions and expectations regarding AI systems. Since these can generate negative perceptions (e.g. distrust in the effectiveness of automated decision making), that can hinder testing and adaptation, it is necessary to develop systematic learning processes based on user feedback to ensure that AI systems are implemented effectively.

References

- [1] A.L. Fogel, and J. C. Kvedar. Artificial intelligence powers digital medicine. *NPJ Digit Med.* **1.1** (2018), 5.
- [2] P. Hamet, and J. Tremblay. Artificial intelligence in medicine. *Metabolism.* **69** (2017), S36-S40.
- [3] B. Mesko. The role of artificial intelligence in precision medicine. *Expert Rev Precis Med Drug Dev.* **2** (2017), 239-241.
- [4] A. Rajkomar, J. Dean, and I. Kohane. Machine learning in medicine. *NEJM.* **380** (2019), 1347-1358.
- [5] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* **542** (2017), 115.
- [6] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans. When will AI exceed human performance? Evidence from AI experts. *J Artif Intell Res.* **62** (2018), 729-754.
- [7] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* **2** (2017), 230-243.
- [8] T.B. Murdoch, and A.S. Detsky. The inevitable application of big data to health care. *JAMA.* **309** (2013), 1351-1352.
- [9] J. He, S.L. Baxter, J. Xu, J. Xu, X. Zhou, and K. Zhang. The practical implementation of artificial intelligence technologies in medicine. *Nat Med.* **25** (2019), 30.
- [10] T. Panch, H. Mattie, and L.A. Celi. The "inconvenient truth" about AI in healthcare. *NPJ Digit Med.* **2** (2019), 1-3. <https://doi.org/10.1038/s41746-019-0155-4>.
- [11] I.G. Cohen, R. Amarasingham, A. Shah, B. Xie, and B. Lo. The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Aff.* **33** (2014), 1139-1147.
- [12] M.J. Rigby. Ethical dimensions of using artificial intelligence in health care. *AMA J Ethics.* **21** (2019), 121-124.
- [13] A. Tang, R. Tam, A. Cadrin-Chênevert, W. Guest, J. Chong, J. Barfett, L. Chepelev, R. Cairns, J.R. Mitchell, M.D. Cicero, and M.G. Poudrette. Canadian Association of Radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J.* **69** (2018), 120-135.
- [14] E.J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* **25** (2019), 44.
- [15] P. Schwartz-Shea, P., Judging quality. In D. Yanow, and P. Schwartz-Shea. eds. *Interpretation and Method: Empirical Research Methods and the Interpretative Turn 2nd Edition*. Routledge, New York, 2015, 120-146.
- [16] A. Langley. Strategies for theorizing from process data. *Acad Manage Rev.* **24** (1999), 691-710.
- [17] M. B. Miles, A.M. Huberman, *Qualitative Data Analysis An Expanded Sourcebook*. Sage Publications, Thousand Oaks, CA, 1994.

Learning Interpretable Behavioral Engagement for Care Management

Subhro DAS^{a,1}, Chandramouli MADURI^b, Ching-Hua CHEN^b and
Pei-Yun S. HSUEH^b

^aMIT-IBM Watson AI Lab, IBM Research, USA

^bIBM Thomas J. Watson Research Center, USA

Abstract. The health outcomes of high-need patients can be substantially influenced by the degree of patient engagement in their own care. The role of care managers (CMs) includes enrolling patients and keeping them sufficiently engaged in care programs, so that patients complete assigned goals leading to improvement in their health outcomes. Here, we present a data-driven behavioral engagement scoring (BES) pipeline that can compute the patients' engagement level with regards to their interest in: (1) enrolling into a relevant care program, and (2) completing program goals. This score is leveraged to predict a patient's propensity to respond to CMs' actions. Using real-world care management data, we show that the BES pipeline successfully predicts patient engagement and provides interpretable insights to CMs, using prototypical patient cases as a point of reference, without sacrificing prediction performance.

Keywords. Care management, personalization, metric learning, interpretability

1. Introduction

Care management is a patient-centered approach to population health that is “designed to assist patients and their support systems in managing medical conditions more effectively.” The Agency for Healthcare Research and Quality (AHRQ) recommends [1] that researchers should investigate (a) the benefit of care management services to different patient segments and (b) the parameters that affect modifiable risks. This can be especially important for those with chronic conditions and those who are transitioning to home care after being discharged from hospital [2].

While machine learning (ML) models may be helpful for patient segmentation and risk analysis, there is a desire for such models to be more explainable, or interpretable, so that they are more likely to be acted upon by human decision makers [3]. Recent reviews [4] have shown that a majority of studies in model interpretability are tied to the optimization of certain model properties for extracting explainable insights. While prior studies have attempted to identify risk stratification and disease progression parameters that affect modifiable clinical risks [5], to our knowledge, this is the first study that has proposed to learn engagement strategies directly from data, based on the quantifiable difference of patient engagement levels and segment-differentiating parameters that affect modifiable risk factors beyond those related to clinical risks.

¹Corresponding Author: Subhro Das, 75 Binney St, Cambridge, MA 02142, USA; subhro.das@ibm.com.

Table 1. Goal attainment records of patients distributed across focus areas & intervention categories.

Focus Areas	Coaching	Coordination	Education	Referral	Screening	Tracking	Other	Total	Status Met
Educational	138	18	277	100	65	0	23	621	81.62%
Implementation	3	192	4	7	0	0	0	206	98.54%
Medications	7	96	30	7	90	0	10	240	84.91%
Reducing Risks	0	1	0	545	53	0	1	600	97.00%
Self-care	130	0	189	12	116	0	49	496	78.69%
Other	29	0	2561	14	29	12	19	2644	99.19%
Total	307	307	3061	685	353	12	102	4827	

This paper focuses on using ML methods to identify patient segments who are most likely to engage in care management, and on care manager (CM) decision support for explaining why they may or may not engage. While a few prior studies have measured patient engagement via quantitative surveys [6], no prior studies have successfully shown to measure patient engagement from real-world data. Its success will be critical for informing individual-level intervention decisions for care management [7].

2. Data

Care Management Records: The care management data, based on which we developed our method, are phone conversation logs of a private, not-for-profit network of community and specialty hospitals. It includes 4,504 transition of care and 440 chronic care related interactions of 3693 unique patients over a 22-month period (Dec-2015 to Oct-2017). From each patient’s anonymized (by Mongo Patient ID) engagement interaction history, we extracted 53 features that includes: (a) basic demographic information (e.g., age, gender); (b) patient’s care program context (e.g., program experience, whether the patient is enrolled, number of days in the program etc); and, (c) interactional context between CMs and patients (e.g., the day of the week when the recorded call occurred). Using this data, we aim to introduce a data-driven approach to achieve two crucial tasks for decision support: program enrollment (“ENROLL”) and goal attainment (“GOAL”). The ENROLL dataset captures the enrollment status for the two assigned care programs: “Transition” include those transitioning from hospital to home after being discharged, and “Involve” include those involving with chronic disease (diabetes etc.) management.

The GOAL dataset, composed of 28 different goals, is classified into six focus areas based on their similarity: Educational (e.g., demonstrate understanding of post discharge, diabetes education), Implementation (e.g., adequate functional, transportation, support for healthy coping), Medications (e.g., adhere with medication regimen), Reducing Risks (e.g., resolve care gaps), Self Care (e.g., understand the benefits of being physically active & healthy diet needs), and Other (e.g., effective care transition). Table 1 summarizes the goal attainment rate for each goal focus area, i.e., the ratio of the number of goals recorded as ‘met’ to the total number of goals in each focus area. The interventions of each goal area are grouped into seven categories: Referral (e.g., refer to a nutritionist for diabetes nutrition education), Education (e.g., educate patients on the importance of physical activity), Coordination (e.g., follow up with providers on refills), Screening (e.g., assess breathing symptoms), Coaching (e.g., provide a log for side effect recording), and Other (e.g., follow up with provider treatment).

3. Methods: Behavioral Engagement Scoring Pipeline

To further incorporate the data-driven approach for indicating patient engagement level and producing interpretable insights, in this paper, we introduce the Behavioral Engagement Scoring (BES) pipeline, which is composed of three components:

3.a) Engagement Outcome-driven Distance Learning: This component aims to learn an engagement outcome-driven distance metric so as to project patient feature-based vectors onto a subspace wherein patients with similar outcomes are closer to each other. Such metrics would be essential to ensure those with opposite outcomes are far away from each other. Here, we adapt Locally Supervised Metric Learner (LSML) [8], which helps estimate the outcome-adjusted distances among patients in the newly transformed vector subspace. The patient features are represented as $X \in \Omega$, where Ω is the vector space, and the class labels as $Y \in \{0, 1\}$. For program enrollment, $y = 1$, if the status is "Enrolled" or "Completed", and, $y = 0$, if the status is "DidNotEnroll" or "Disenrolled". Similarly, for goal attainment, $y = 1$, if the status is "Met", and, $y = 0$, otherwise.

We minimize the generalized Mahalanobis distance across all CM records segmented into homogeneous (records with same outcome) and heterogeneous (records with opposite outcomes) neighborhoods. The feature transformation matrix that minimizes this generalized Mahalanobis distance, renders the data into the desired space, where records with similar outcomes are compact and those with opposite outcomes are distant. Refer to [8] for the complete LSML algorithm that derives this feature transformation matrix. We employ this matrix to obtain the projected feature set. For the rest of the pipeline, we leverage these outcome-adjusted projected features and the corresponding Mahalanobis distance between patient-based vectors to learn patient segments and to estimate each patient's propensity to engage with CMs.

3.b) Learning Patient Segment-based Behavioral Profiles: This component aims to capture patient segments wherein patients and CMs interact with a similar behavioral profile. However, the optimal number of patient segments is not known "a priori". With that objective and constraint in mind, hierarchical clustering [9] is employed to identify patient segments based on the outcome-adjusted distances in the newly projected vector subspace and to learn the key factors that drive differential engagement outcomes. We choose Complete Linkage method (as it has highest outcome-differentiation power, measured by ANOVA scores) to compute inter-segment similarity of the furthest pairs.

To compute the optimal number of segments, Elbow method (an automatic tuning algorithm) tracks the acceleration of distance growth among segments and thresholds the agglomeration at the point where the acceleration is the highest. Hence, our population is clustered into k segments, each capturing distinctive patterns that drive differential patient responses in engagement. We further identify prototypical patient cases in each segment to serve as examples to showcase the distinctive patterns in their behavioral profile and to add interpretability to the engagement scores produced by the BES pipeline. The prototypical patient cases are those with positive engagement outcome, i.e., $y = 1$, while being closest to the centroid of each patient segment in the transformed feature space.

3.c) Estimating propensity to respond: In this component, we train a Generalized Linear Model (GLM) to compute the engagement score for patients who have been assigned a program to enroll or a goal to attain, using the behavioral segment information found in 3.b. The GLM feature weights for each segment k are computed by minimizing the least squared errors over all the data points from that segment k . This trained GLM with optimized feature weights is used to compute the scores for new patients, which represents their propensity to respond to CM's engagement calls or intervention assignments for goal completion. The last stage of our pipeline contains a Support Vector Machine (SVM) to classify the potential engagement outcome of each patient using their BES.

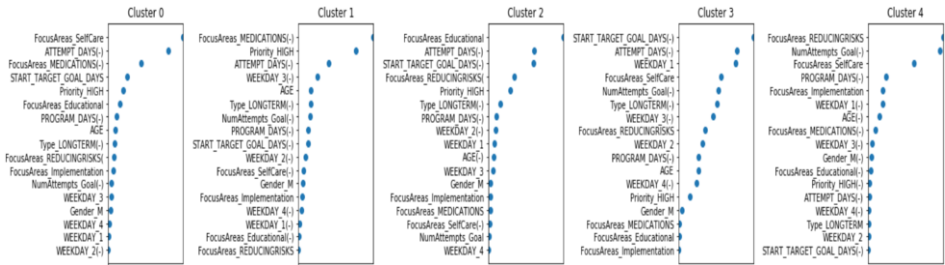


Fig. 1. Feature weights indicating patient response driver across 5 behavioral segments for goal attainment.

4. Results & Discussion

The BES pipeline, introduced in this paper, segments patients based on not only the patterns exhibited during patient-CM interactions, but also their engagement outcomes. Although each feature contains only weak signals to differentiate overall engagement outcomes, when considered collectively in a segment, the combined feature sets can explain rich engagement behaviors for care planning.

Performance evaluation on engagement outcome prediction: The predicted engagement outcome of each patient task is compared with what actually happened as indicated in the CM transaction records. The results across all patient tasks are aggregated to evaluate the overall performance of the BES pipeline on engagement outcome prediction in terms of precision, recall, accuracy and F1-score. In particular, the evaluation aims to compare the following three versions. The first “Population-based (Baseline)” directly trains a SVM classifier in the BES pipeline using all the population data without differentiating patient segments. The second “Behavioral Profile-Driven” trains BES for each of the patient segments using data belonging to patients in that segment. The third “Prototypical User-Driven” trains BES using only prototypical patient cases in each segment. Table 2 shows the performance evaluated with 5-fold cross validation, indicating that the BES pipeline helps predict patient responses (“whether to engage”) for each type of engagement tasks with high precision (> 90%). This implies a high percentage of successful engagements if BES-based recommendations are followed for task prioritization.

Drivers of Differential Patient Response: We analyze feature weights of the model trained for each of the five patient segments in order to pinpoint drivers that contribute the most to its engagement outcome prediction. The results are used to provide support for CMs to understand the rationale of the predictions. Fig 1 demonstrates that feature rankings are significantly different when the population-level feature rankings are compared with the segment-based ones (Spearman’s coefficient ρ ; $p < 0.01$). In some segments,

Table 2. Performance with 5-fold cross-validation.

Method	Performance Metrics	Program Enrollment: Involve	Program Enrollment: Involve	Goal Attainment
Population Based (BASELINE)	Accuracy	0.96	0.96	0.89
	Precision	0.95	0.94	0.93
	Recall	0.99	1	0.95
	F1	0.97	0.97	0.94
Behavior Profile-Driven	Accuracy	0.93	0.74	0.79
	Precision	0.95	0.94	0.92
	Recall	0.96	0.65	0.83
	F1	0.95	0.76	0.87
Prototypical User-Driven	Accuracy	0.93	0.74	0.95
	Precision	0.95	0.93	0.99
	Recall	0.96	0.65	0.96
	F1	0.95	0.76	0.97

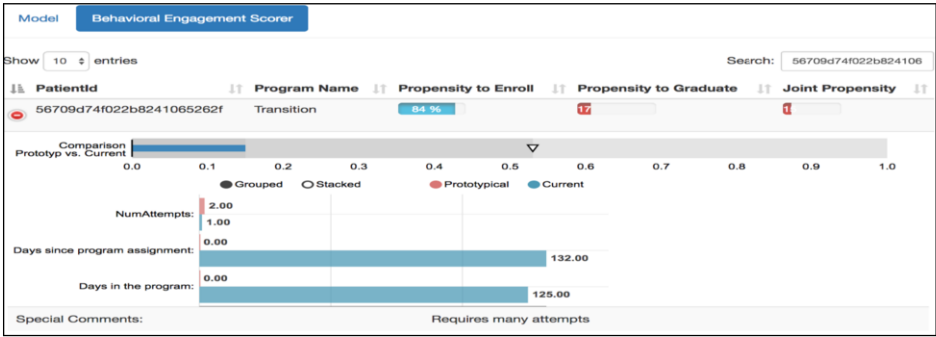


Fig. 2. Interpretable engagement insights for Care Managers for shared decision making.

we observe strong indicators of CM influence (e.g., how long a CM has been trying to help the target patient obtain his/her goal, how many calls have been attempted before the goal attainment status is determined). In other patient segments, we observe certain goals as more likely to yield positive engagement responses.

Interface and API for Shared Decision Making: We developed a web-based user interface that provides the CMs access to explainable insights about the gauged engagement level, the predicted response and the prototypical patient cases. Fig 2 shows the interactive tooling based on a demo API customized for CMs to make their decisions before the call regarding which patients to call first for program enrollment and for goal attainment, and what are the reasons that they might or might not be engaged.

5. Conclusion

The main contributions are: (1) developing a quantitative and personalized approach for identifying patients who are more likely to engage in care management; (2) demonstrating empirically, using real-world data, that our approach provides more accurate engagement outcome predictions compared to a “one-size-fits-all” population-based approach; (3) generating explainable insights on each patient’s behavioral profile from the prototypical patient cases within his/her segment. Performance evaluation results show that, in our case, explainability does not come at the expense of model performance.

References

[1] *Implications for medical practice, health policy, and health services research*. Agency for Healthcare Research and Quality. 2015.

[2] R. Brown, D. Peikes, G. Peterson, J. Schore, and C. Razafindrakoto. Six features of medicare coordinated care demonstration program that cut hospital admission of high-risk patient. *Health Affairs* (2012).

[3] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad. Intelligible Models for Health Care. *ACM KDD Conference Proceedings*, 2015.

[4] P.-Y. Hsueh, S. Dey, S. Das, and T. Wetter. Making sense of patient-generated health data for interpretable patient-centered care: The transition from more to better. *Proceedings of MedInfo*, 2017.

[5] B. Liu, Y. Li, Z. Sun, S. Ghosh, and K. Ng. Early Prediction of Diabetes Complications from Electronic Health Records: A Multi-Task Survival Analysis Approach. *Proceedings of AAAI Conference*, 2018.

[6] G. Graffigna, S. Barello, A. Bonanomi, and E. Lozza. Measuring patient engagement: development and psychometric properties of the patient health engagement (PHE) scale. *Frontiers in psychology* (2015).

[7] P.-Y. Hsueh, S. Das, C. Maduri, and K. Kelly. Learning to personalize from practice: A real world evidence approach of care plan personalization based on differential patient behavioral responses in care management records. *AMIA Annual Symposium Proceedings*, 2018.

[8] J. Sun, F. Wang, J. Hu, and S. Edabollahi. Supervised patient similarity measure of heterogeneous patient records. *ACM SIGKDD Explorations Newsletter* (2012).

[9] J. Friedman, T. Hastie, and R. Tibshirani. The elements of statistical learning, *Springer*, 2001.

Mi Argentina/Mi Salud: The Argentinian Citizen Digital Health Portal

Daniel A. RIZZATO LEDE^a Federico A. PEDERNERA^a Emiliano LÓPEZ^a
Cintia D SPERANZA^a Jesse J. MAID^a Fernando GASSINO^{a,c} Osvaldo ARGIBAY^a
Rodrigo CASTIGLIONE^a Guillermo PUJADAS^a Emanuel GONZALEZ ALVAREDO^a
Rodrigo ÁLVAREZ^a Maria B. MARKIEWICZ^a Florencia ALEGRE^a Leandro
FERRAREIS^a Florencia ROLANDI^b Fabiana AYALA^b Daniel A. ABADIE^b Cintia A.
CEJAS^a Alejandro LÓPEZ OSORNIO^a and Adolfo RUBINSTEIN^a
^a *Ministry of Health, Argentina*
^b *Ministry of Modernisation, Argentina*
^c *Hospital Italiano de Buenos Aires, Argentina*

Abstract. The Ministry of Health (MoH) stated the National Digital Health Strategy 2018-2024 in order to establish the conceptual guidelines for the design and development of interoperable health information systems. It included the creation of a National Digital Health Network, and a Citizen Health Portal to inform and empower patients about their rights. For instance, the Digital Vaccination Card is already available and has equal legal validity as its paper version. The platform also works as a personal privacy manager, to configure the consent for Health Information Exchange through the network, or to check the access logs. This paper outlines the implementation experience of this powerful tool at a national level.

Keywords. eHealth, Personal Health Records, Health Information Exchange, Privacy.

1. Introduction

The development of sustainable Health Systems that promote universal access to care can only be enabled by the appropriate use of digital technologies[1]. According to the Argentine legislation, the medical record can be paper or magnetic supported, and should always be subjected to the principles of integrity, uniqueness, inviolability and confidentiality. There is a legal obligation for the health institutions "...upon request of the patient, owner of the medical record, an authenticated copy must be provided, within a period of forty-eight hours from request...".

Argentina has a complex federal, segmented and highly fragmented healthcare system, which also translates into the generation of information silos. In many health centres around the country, health records are still kept in manual registries and logbooks. In a preliminary survey for the public sector, at least 80% of the healthcare facilities had no computers or internet connection. Therefore, patient's right to access their own information is frequently hard to accomplish.

During the previous 15 years several management systems were implemented at the Ministry of Health (MoH) for the Essential Public Health Functions. Most of them

were independent collecting tools that required paper form fulfilment for later electronic transcription, fuelling the administrative burden. Nevertheless, these systems empowered the MoH with important information for decision making: Surveillance, Immunization, Vital Statistics, Disease Registries, Public Health Insurances, among others. Most of this information is not available to general practitioners or patients themselves.

The Argentinian National Digital Health Strategy 2018-2024[2] has the vision of reducing quality gaps in health care, implementing information systems that identify the characteristics and needs of the population, allowing longitudinal and comprehensive monitoring of people throughout the entire health system and providing innovative tools to health professionals and patients. It encourages every healthcare institution in the country to progressively adopt any interoperable electronic health record (EHR) system. The National Digital Health Network was launched in April 2019, using an enterprise service bus technological infrastructure to connect different systems across the country. The data remains in each information system repository, and is only communicated to other domains if necessary (e.g. when the patient visit another clinic), and the citizen has given his consent for Health Information Exchange.

A Personal Health Record (PHR) has been defined as a set of computer-based tools that allow people to access and coordinate their lifelong health information and make appropriate parts of it available to those who need it. They usually consist of provider-tethered applications that allow patients to electronically access health information documented and managed by a healthcare institution. Although patient portals are already being implemented, it is still unclear in which ways these technologies can influence patient care[3]. In digitally advanced countries like Estonia, every person has a personal account in a patient portal, where they can review all their healthcare data and insurance expenses[4]. My Health Record is Australia's electronic PHR system, which from 2012 to 2017 got 21% of registered population. There were low levels of awareness and concerns about sharing records and equity of access for all Australians[5]. The government changed patient enrolment from opt in to opt out, reaching 90% of the population by 2019. There are not many international examples of Vaccination eCards, despite it seems there would be many benefits with its implementation[6].

In 2017 the Argentinian Ministry of Modernisation (MoM) jointly with the MoH launched a Citizen Digital Health Portal (*Mi Argentina / Mi Salud*) for personal access, empowering the people and recognizing their pivotal role in their own health. This paper describes the implementation of the Citizen Health Portal in Argentina.

2. Materials and Methods

2.1. Setting

Argentina is a large country of more than 1 million square miles located in the southern half of South America. Its population is around 45 million people. The sovereign state is a Federal Republic subdivided into twenty-three provinces and one autonomous city. Beyond the recurrent economic crises, it is an upper middle-income country, and has a developed healthcare system. It has a quite complex organisation, with 3 subsystems: Public Care, Social Security and Private Insurances. Many of its shortcomings come from its segmented and highly fragmented nature. Since 2018, the MoH has embraced

an action plan to implement a public health reform to provide actual Universal Health Coverage (UHC) by strengthening provincial health systems[7].

Digital Health is one of the three pillars for the effective implementation of UHC. Through Resolution 189/2018, the highest health authority in Argentina settled the way forward to build, in collaboration with provincial governments and referents of the 3 subsystems, a network of information systems focused both on the population and individual health of each patient.

2.2. Methods

The Citizen Digital Health Portal was designed and developed following the user centered design and accessibility rules promoted by the MoM[8]. People should be able to access their health data in an easy and comprehensive way[9]. Usability was tested, but it is not in the scope of this paper. The different modules of *Mi Argentina* -the Citizen Digital Platform- are described, including *Mi Salud*, the health section. We assessed its utilisation with quantitative aggregated and de-identified measurements of different services provided by this tool.

3. Results

The Citizen Digital Health Portal *Mi Salud* is the health section of *Mi Argentina*[10], the digital public service delivery platform developed by the MoM. In 2017 the Argentinean Government unified the official landing website in the Argentina.gob.ar domain. These two platforms are crucial milestones in terms of digital service delivery in the country. Important work has been undertaken, following the example of the United Kingdom (gov.uk), and other OECD members and partner countries in the region like Mexico (gob.mx) and Peru (gob.pe)[11]. As a result, the government-citizen relationship builds on the use of Argentina.gob.ar website and *Mi Argentina* platform. Both were designed and implemented by the Service Delivery Team within the MoM. Unique personal identification, privacy and security were undertaken by different modalities of accreditation, the highest one requiring biometric validation. Since its deployment in 2017, the number of users has been constantly growing. As of October 2019, more than 2.3 million users had registered on *Mi Argentina*. A line chart with the number of active accounts can be seen in Figure 1.

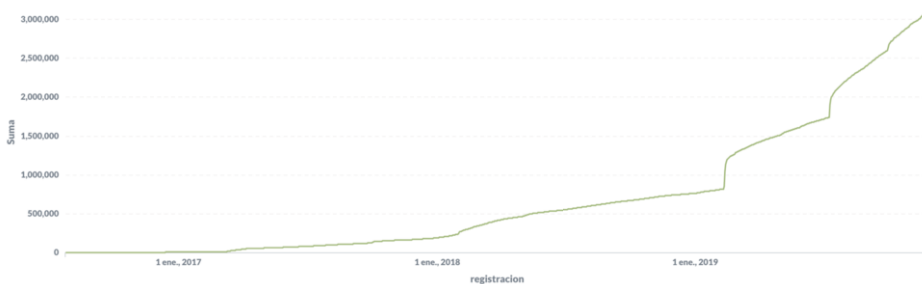


Figure 1. Line chart showing the number of *Mi Argentina* accounts during the last 2 years.

To gain full access to *Mi Argentina*, each citizen must create an account and validate her digital identity through biometric technology. Although this strategy potentially reaches the entire population, it was initially available for citizens over 13 years old who could individually open an account in *Mi Argentina*. Children will be included soon, with special permission from their parents or tutors.

Mi Argentina provides appointment booking (*Turnos*) for government services like immunization and online certifications from the National Administration of Social Services (ANSES) and other offices. There is a Wallet (*Mi Billetera*) where users can access their national identity eCard, driving licences, car documents and insurances, and other digital certificates in one place. *Mi Argentina* can be accessed both web and through mobile application. Screenshots of the App can be seen in Figure 2.

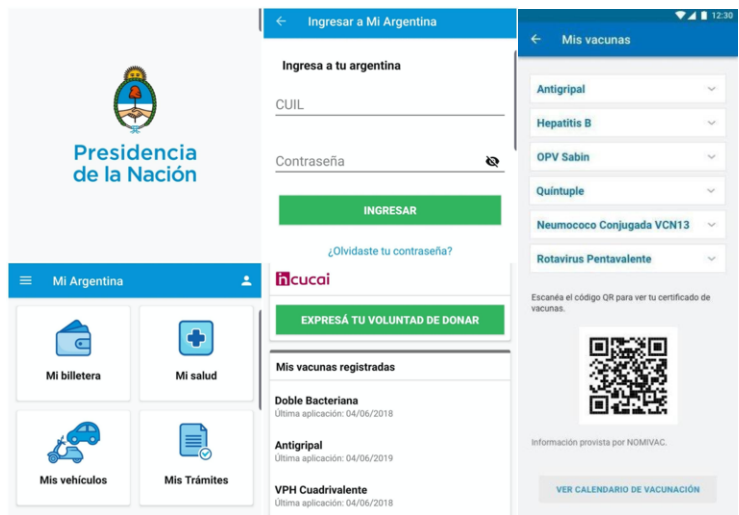


Figure 2. Screenshots of *Mi Argentina* App and *Mi Salud*, the Citizen Digital Health Portal.

The Health Section (*Mi Salud*) includes different services for citizens: a map with the location of healthcare centres, institutions that perform free HIV test, online consent for organ donation will, online access to health formalities like getting the unique disability certificate, and information about the current healthcare coverage. One of the most important milestones was the Health Information Exchange (HIE) consent. The portal itself is a domain of the Digital Health Network. It works as a personal privacy manager, to set the consent for information exchange through the network. Patients can also check the EHR access logs. Up to December 2019, there have been almost 41,000 citizens (73% male) who have set their consents, with 94% of affirmative responses, as seen in Table 1. Even though it could be seen as a low enrolment rate, HIE is based on an opt-out modality.

Table 1. Data regarding patient utilization of the Digital Health Network and the Citizen Portal.

Number of Citizens (2019)	44,938,712
with an EHR linked to the Digital Health Network	2,068,448
with an active <i>Mi Argentina</i> account	2,375,949
with a validated <i>Mi Argentina</i> account (biometric)	922,480
with active consents for Health Information Exchange	40,940 (94% affirmative)

Recently, we implemented in *Mi Argentina* the Unified Digital Vaccination eCard. It is digitally signed and has a QR validation code that gives the same legal validity as the usual cardboard immunization card issued by health institutions. Through the citizen digital platform *Mi Argentina* people can individually see their registered vaccines. There are almost 30,000 daily accesses to the eCard. Patients can report if there is missing data in the digital card. This eCard is built using the data from the Federal Registry of Nominalized Vaccination (NOMIVAC)[12]. NOMIVAC is a module of the Argentinian Public Health Information Management System (SISA) that collects the vaccination data of every citizen in the country. There are more than 52 million registered doses. This module allows any vaccinator to register the applied doses in real time, through the web or a mobile application that even works offline. It also gets data from local systems through interoperable web services, mainly from public centres but progressively also from social security and private healthcare institutions.

4. Conclusion

Even though single institution PHR might not change the traditional passive patient role, we strongly believe that our National Digital Health Portal has the potential of empowering citizens. It allows patients to access their information and change their HIE consent for the National Digital Health Network. The progressive addition of new tools within *Mi Argentina* (Vaccination eCard and soon International Patient Summary, Lab results, ePrescriptions) would enhance such vision. Further analysis of the platform utilisation is needed. Improved promotion and support might lead to higher citizen engagement with *Mi Salud*, especially for populations at risk of digital exclusion.

References

- [1] WHO, and ITU, National eHealth Strategy Toolkit, Geneva, 2012.
- [2] Argentinian Digital Health Strategy 2018-2024, (n.d.). <https://argentina.gob.ar/salud/digital>.
- [3] E. Ammenwerth, P. Schnell-Inderst, and A. Hoerbst, The impact of electronic patient portals on patient care: a systematic review of controlled trials., *J. Med. Internet Res.* **14** (2012) e162.
- [4] E. Bastias-Butler and A. Ulrich, Digital Transformation of the Health Sector in Latin America and the Caribbean: Electronic Health Records., *Social Protection and Health Division Regional Policy Dialogue Report*, Interamerican Development Bank, Washington DC, USA, 2019.
- [5] Y. Van Kasteren, A. Maeder, P.A. Williams, and R. Damarell, Consumer Perspectives on MyHealth Record: A Review, *Stud Health Technol Inform.* **239** (2017)146-152.
- [6] K. Wilson, K.M. Atkinson, and C.P. Bell, Travel Vaccines Enter the Digital Age: Creating a Virtual Immunization Record, *Am J Trop Med Hyg.* **3**(2016): 485–488. doi: 10.4269/ajtmh.15-0510
- [7] A. Rubinstein, M.C. Zerbino, C. Cejas, and A. López, Making Universal Health Care Effective in Argentina : A Blueprint for Reform, *Heal. Syst. Reform.* **4** (2018) 203–213.
- [8] Web Accesibility in Argentina, (n.d.). <https://www.argentina.gob.ar/onti/accesibilidad-web>.
- [9] M.I. Smith, G. Garcia, M. Simon, L. Bruchanski, J. Frangella, J.A. Sommer, M. V Giussi, and D.R. Luna, Lessons Learned After Redesigning a Personal Health Record., *Stud. Health Technol. Inform.* **245** (2017) 216–220.
- [10] Mi Argentina, (n.d.). <https://mi.argentina.gob.ar/>.
- [11] OECD, Digital Government Review of Argentina: Accelerating the Digitalisation of the Public Sector, Paris, 2018.
- [12] D.A. Rizzato Ledo, and Z.A. González, *Sistemas de información para la gestión sanitaria, 1st ed.*, Hospital Italiano de Buenos Aires, Argentina, 2018.

Nutrikids, a Smartphone Application to Improve the Quality of Paediatric Dietary Assessments : A Feasibility Study

Léa JACQUES^{a,1}, Catherine BUSSIEN^b, Camille DESCLOUX^c, Lisa DECORSAIRE^d, Christian LOVIS^{a,b}, Frédéric EHRLER^b

^a Faculty of Medicine, Geneva University, Switzerland

^b Division of medical information Sciences, Geneva University Hospitals, Switzerland

^c Haute école de santé (Geneva HEDS), Switzerland

^d Grenoble Polytechnic School, France

Abstract. Collecting accurate food consumption data is challenging for nutrition practitioners. Mobile devices are interesting alternatives to traditional ways of collecting assessments by reducing burden of use. Our objective was to develop and evaluate a mobile application for food intake documentation, targeting particularly the parents of pediatric patients. NutriKids is an application designed to guide users step by step in their assessment by asking relevant questions and taking advantage of the camera function. The quality of nutritional assessment using NutriKids was evaluated through the analysis of food diaries completed during two days under real-life conditions (home use). NutriKids demonstrated to provide satisfying degree of quality in 84% of the reported items. A detailed analysis of quality components highlighted that NutriKids interface could be improved for quantitative estimation and for specific information on added fat or sugar. Besides, a feedback survey highlighted the high acceptability of the application by the participants. These findings demonstrate the potential of this guided approach application. With further development of NutriKids and controlled based assessment with larger populations, we hypothesize that our results would reach statistical significance in all aspects of nutritional assessment quality.

Keywords. mHealth, mobile applications, nutrition assessment, diet records, child

1. Introduction

In the pediatric setting, dietary assessments are required in a wide range of medical conditions, such as children with weight problems, gastrointestinal disorders, metabolic or endocrine disorders, etc. Collecting accurate food and beverage consumption data from patients is challenging for nutrition practitioners. Thus far, three methods have mainly been used: Food Frequency Questionnaires (FFQ), 24H Dietary Recalls (24HR) and Food Records [1]. These methods usually require parents to write down precisely everything eaten and drunk during a few days, using different means such as

¹ Mrs. Lea Jacques, Geneva University, Faculty of Medicine; E-mail: lea.jacques@etu.unige.ch

questionnaires or paper journals in the form of a blank table. This task requires personal investment and time. Moreover, traditional paper tools are hampered by practical limitations, which generate burden for the user and thus challenges the cooperation between user and practitioner.

With digitalization, new dietary assessment methods are developing rapidly. Comparative studies have already demonstrated the positive impact of digital technologies over traditional paper-based methods by increasing participation rates and reducing the burden of use [2]. However, evidence validating their efficiency is still lacking and the accuracy of collected data, namely precise description and quantitative estimation of the reported food, remains a critical point.

Digital tools to quantify food servings have greatly evolved over time. Early versions help users by displaying different serving sizes pictures. Studies have demonstrated that these tools were associated with lower precision compared to weighing [3], although they increased accuracy of intake estimation when no scale is available [4]. Other techniques rely on photographs taken by the user which are then analyzed by dietitians. However, in studies evaluating dietitians' proficiency to estimate dietary intake based on digital food images (pictures of meals), it appears that the quantification accuracy is quite low without specific training for this task [5,6]. More recently, new approaches relying on artificial intelligence, featuring automatic food identification and portion size estimation are under development (e.g. NutriNet, Snap-n-Eat). Regarding the precision of food description, studies have shown that long exhaustive lists of food items based on food databases don't improve quality. Overall, existing tools still suffer from suboptimal reliability, and tools based on guided approach may represent a promising improvement.

1.1. NutriKids

We developed a new mobile application, NutriKids, in partnership with a group of experts in paediatric dietetics at the Geneva Children's Hospital. NutriKids is meant to offer a smartphone alternative to actual paper journals (food record method). The creation process started by analyzing recent nutritional assessments to identify which information should be collected. It highlighted the importance of three elements for assessing valuable dietary intake, namely description, quantity and information on added fat/sugar. From there, we designed our application. The aim was to guide parents in the assessment process in order to minimize the omission of key information previously identified. For example, no food item can be registered if its quantity hasn't been indicated. More specifically, for every food item potentially cooked or seasoned with fat, the application automatically asks the user about the amount of added fat (see fig. 1). A camera function is also provided to capture food labels or product specificities, reducing the burden of writing down descriptions. Besides, we set up a system of « typical meals »: when the users select the main food item they wish to report, other items, expected to be associated with the main one, are automatically proposed (see fig. 1). We decided to avoid long exhaustive lists of food items that users can select. Instead, we preferred short lists containing frequently selected items and added the possibility for the user to write down his own description or to take pictures of the product. These particularities differentiate NutriKids from the numerous existing applications that we have tested on the market (e.g. Gluci-Check, MealLogger, SHYE, Food(lg), FatSecret, iEatWell), which are already very useful for some functions but lack the guiding aspect in data collection.

After the implementation of the first version, we performed an expert evaluation using the Nielsen criteria [7] and modified few elements to enhance the application's coherence and efficiency of use.

A usability test, performed by a distinct group of 10 users with the application's final version, confirmed that users don't encounter major difficulties using the application.

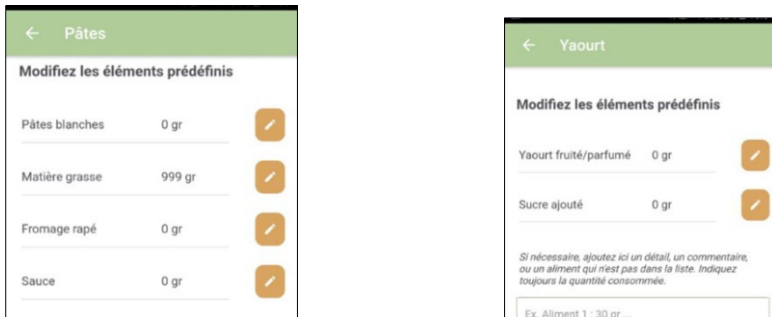


Figure 1. Screenshots of NutriKids' interface displaying the "typical meal" structure as well as the automatic recall for fat/sugar quantity

Following the creation of the application, we evaluated NutriKids during a two-day assessment period in a pediatric setting. This period of time was set to be very close to the current conditions of dietary assessment at the hospital (3 days), to avoid bias of motivation.

2. Method

2.1. Evaluation

We performed a critical evaluation of the quality of the data reported by a group of 12 participants during a dietary assessment. Participants included parents from pediatric patients requiring a nutritional assessment or having completed such an assessment in the past two years, as well as volunteering parents from the local community.

Participants were requested to report all food and beverage intake within a period of two days using NutriKids. Once collected and anonymized by the investigator, data were then analyzed by hospital dietitians according to the criteria identified as being essential for a valuable dietary assessment:

- Information about added fat and/or sugar (if relevant)
- Detailed description of ingredients and/or picture of food brand
- Precise weight (given in gr., ml., tsp. or Tbsp.) or picture of the portion

For every food or drink item reported during the two-day evaluation period, each criterion was evaluated by a dietitian as « sufficient » or « insufficient », depending on whether additional information was required or not in order to evaluate properly the nutritional content of the meal. Thus, on the two-day period, we calculated for each criterion (description, quantity and added fat/sugar) the percentage of items reported correctly: 0% being no reporting deemed sufficient at all and 100% being perfect reporting for all items. Each criterion had identical ponderation for the evaluation of the

global quality (primary outcome), which was calculated as the mean of the three criteria.

We calculated the group results as the mean of participants' individual score.

To limit assessment biases, a second evaluator independently controlled 10% of the data and found no significant discordance (kappa analysis > 0.8).

At the end of the evaluation period, participants also had to complete a survey containing Likert based questions, as well as open questions providing a more qualitative evaluation of the application.

3. Results

3.1. Evaluation of NutriKids

All 12 participants were women, mostly aged between 35 and 45 years and from higher educational levels. 88% of them were familiar with smartphone use. 55% had medical reasons to participate, 44% volunteered.

The total number of food and drink items reported for all 12 participants was 193, among which 69 needed information about added fat or sugar. As shown in table 1, mean global reporting quality was rated as 83.80%. Quality was specifically considered as "sufficient" for descriptions in 95.52%, for quantitative estimation in 88.20% and for added fat/sugar information in 67.68%.

Table 1. Mean quality of the assessment for each criterion

	n	Mean (%) ± SD
Global quality		83.80 ± 15.66
Food description	185 (N=193)	95.52 ± 6.86
Quantitative estimation	173 (N=193)	88.20 ± 14.62
Added fat/sugar information	50 (N=69)	67.68% ± 33.95

3.2. Feedback evaluation form

Participants rated the general ease of use of NutriKids 7.4/10 and its intuitiveness 8.9/10. The major difficulties reported were about finding desired food items in the lists and informing about food's quantity, especially when done without using a scale or when the meal was complex. The majority of the participants used the camera function and reported it to be very useful.

Interesting improvement suggestions concerned the possibility to add notes with recipes, or the possibility to create a personal list of specific items to reuse, avoiding the burden of copying all information again if often consumed.

4. Conclusion

The evaluation of NutriKids confirms previous findings that mobile applications may represent a valid alternative to traditional paper-based methods. The quality of intake assessment was globally good, and participants were mostly satisfied with NutriKids.

Separate analysis of each quality criterion underscores which aspects cause difficulties to participants and could benefit from further improvement. Precise

description of food does not appear to be a problem given that more than 90% of the items are evaluated as sufficiently well described. Difficulties were mostly encountered for quantitative estimation, especially when no scale is at hand. Possible improvements include image-assisted tools such as a food atlas (ranges of food photographs displaying different serving sizes with associated quantities) or machine learning (real-time analysis for automatic food recognition and weight estimation). However, even though machine learning has demonstrated its potential in improving quantitative estimation and reducing underreported food [8,9], it may require a certain time before being used out of the lab and may not solve the problem of complex meals with mixed ingredients.

Concerning information on added fat or sugar, our fairly good results for this specific aspect of food reporting strengthen our hypothesis that a guided approach represents a valuable tool to improve the quality of nutritional assessments. To our knowledge, no other study has addressed the issue of a guided approach up to now.

4.1. Improvements, Perspective

NutriKids' interface could be improved to guide even more the food collecting process. Users could, for instance, be constrained to inform about added fat or more clearly invited to provide pictures. Additional research should be undertaken to find the best solution for quantitative estimation when no scale is available. No clear solution has yet been found to achieve this error prone task, but a combination of different image-assisted methods could be sufficient for daily practice. With those few adjustments, we hope to provide a valid alternative to the traditional paper journal.

References

- [1] F.E. Thompson, A.F. Subar, C.M. Loria, J.L. Reedy, T. Baranowski, Need for technological innovation in dietary assessment, *J Am Diet Assoc.* **110** (2010), 48-51.
- [2] Traditional methods v. new technologies – dilemmas for dietary assessment in large-scale nutrition surveys and studies: a report following an international panel discussion at the 9th International Conference on Diet and Activity Methods (ICDAM9), Brisbane, 3 September 2015.
- [3] E. Foster, A. Hawkins, K.L. Barton, E. Stamp, J.N.S Matthews, A.J Adamson, Development of food photographs for use with children aged 18 months to 16 years: Comparison against weighed food diaries – The Young Person's Food Atlas (UK), *PLoS One* **12(2)** (2017), e0169084.
- [4] M. Nelson, M. Atkinson, S. Darbyshire, Food photography II: use of food photographs for estimating portion size and the nutrient content of meals, *Br. J. Nutr.* **76** (1996), 31-49.
- [5] A.A. Fatehah, B.K. Poh, S.N. Shanita, J.E. Wong, Feasibility of Reviewing Digital Food Images for Dietary Assessment among Nutrition Professionals, *Nutrients* **10(8)** (2018), 984.
- [6] M. Vasiloglou, S. Mougiakakou, E. Aubry, A. Bokelmann, R. Fricker, F. Gomes et al. A Comparative Study on Carbohydrate Estimation: GoCARB vs. Dietitians, *Nutrients* **10** (2018), 741.
- [7] J. Nielsen, Enhancing the explanatory power of usability heuristics. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (1994), 152-158.
- [8] C.J. Boushey, M. Spoden, E.J. Delp, F. Zhu, M. Bosch, Z. Ahmad et al. Reported Energy Intake Accuracy Compared to Doubly Labeled Water and Usability of the Mobile Food Record among Community Dwelling Adults, *Nutrients* **9(3)** (2017), 312.
- [9] C.K. Martin, J.B. Correa, H. Han, H.R. Allen, J.C. Rood, C.M. Champagne et al. Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time, *Obesity (Silver Spring)* **20(4)** (2012), 891-9.

On Patient Accessible Electronic Health Records and the Experienced Effect on the Work Environment of Nurses

Jonas MOLL ^{a,1} and Åsa CAJANDER ^b

^a Centre for empirical research on information systems, Örebro University, Sweden

^b Department of Information Technology, Uppsala University, Sweden

Abstract. The work environment for nurses is challenging, and even though new technology has major advantages, it has often also caused new stressors and problems for nurses. When Patient Accessible Electronic Health Records (PAEHR) was introduced in Sweden, research showed that nurses were worried about possible negative effects for patients as well as on their work. However, to this date there are very few follow up studies on the more long term effects of PAEHR, despite research pointing to the fact that reactions after long term use might differ from initial experiences. In this paper we present an interview study to fill this research gap. An analysis based on interviews with physicians and nurses in Oncology care reveals three areas where nurses' work is experienced to have been affected: 1) nurses receive more questions from patients after PAEHR has been introduced, 2) nurses have changed their documentation practices and 3) the log list functionality has made nurses feel questioned. Finally, these results are discussed in relation to nurses' work environment from a sociotechnical and gender perspective.

Keywords. eHealth, patient accessible electronic health records, work environment, nurse, interview

1. Introduction

The work environment in health care is indeed challenging and even though computer systems have managed to improve some areas and make them more efficient, they have also brought new problems and challenges into nurses' work. Nurses are currently faced with increasing workloads due to patients with co-morbidities, staff shortages and increased administrative work [1] [2] on top of new technology being implemented on a regular basis that has unwanted negative side effects for nurses [3].

When Patient Accessible Electronic Health Records (PAEHR) were launched in Sweden in 2012 nurses were skeptical to the system, which has been shown in a large survey study [4] and a small interview study in primary care [5]. Areas of worry related to the risk of PAEHR creating added work through changed work practises, changes in documentation practices, increased stress and worsened work environment. Other studies also pointed to nurses in psychiatry care worrying about PAEHR contributing to an in-

¹Corresponding Author: Jonas Moll, Örebro University, Fakultetsgatan 1, S-701 82 Örebro, Sweden; E-mail: jonas.moll@oru.se.

creased number of threats and violence due to patients knowing names of nurses, through the access log list, and being able to contact them [6] [7].

A sociotechnical system, such as nursing and PAEHR, cannot be properly understood without taking gender into account. Physicians and nurses do gender [8], i.e. perform what is expected from them as women and men in relation to other people, and in relation to technology. Studies on gender and technology reveal that communication patterns in nursing are influenced by gender and gender power, and that nursing is traditionally seen as a nurturing and emotional work [9]. However, too little is known about how the nursing profession is affected by technology.

Moreover, there are still very few studies looking into the experienced effect on work as PAEHR has reached a more mature use - most earlier studies in the area have been carried out within a year after launch. Initial experiences with and reactions towards new technical solutions, like the PAEHR in this case, are often different from experiences and reactions after more long-term use [10]. Given the current work environment problems in health care, this is a crucial area to look into from a societal perspective, and from a research perspective, more research is needed on the connection between digitalisation and the work environment. Additionally, most studies to date on the effect of PAEHR on the work environment of healthcare professionals have focused on effects for physicians [11] [12] [13] [14]. Hence, this paper looks into the experienced more long term effects of PAEHR on the work environment of nurses - an area which is under-investigated and in need of a qualitative research approach. More specifically the paper looks into the experienced effects on the work of nurses including general effects on workload and documentation and a specific technical part - the access log list.

2. Method

This paper reports on preliminary results from an interview study at Uppsala University Hospital. The study was ethically approved by the Regional Review Board in Uppsala (EPN 2017/045).

The results presented here are based on a content analysis of semi-structured interviews with 15 physicians and five nurses at the Oncology department. The interviewees were recruited through the head of department as well as the department's business developer. The interviews were held in rooms chosen by the interviewees, most often in empty examination rooms or private offices. Each interview took around 45 minutes and covered several topics, such as attitudes towards PAEHR, effects on work and effects on various aspects of communication with patients. The interviews were recorded and transcribed in verbatim prior to the analysis.

This paper focuses on results related to the effects on the work of nurses, including general effects on workload and documentation, and a specific technical part - the access log list - is treated in depth since it was shown to affect nurses in particular.

3. Results

The results describe in what ways PAEHR is experienced to have a clear impact on the Oncology nurses' work.

3.1. Nurses Receive More Questions as They Are the First Line of Support for Patients

According to the majority of the interviewed physicians, nurses in oncology care - and so-called *contact nurses* in particular - act as the first line of contact with patients. Consequently, patients ask nurses more often about health care conditions, test results and procedures. Accordingly, when patients call the hospital to ask for example about the information they have found in the PAEHR, they reach one of the nurses. According to several physicians, this has resulted in nurses receiving a lot of questions that they cannot answer questions that are then forwarded to a physician. An increase in questions from patients and questions that nurses cannot answer in particular - after the launch of the PAEHR is also confirmed in the interviews conducted with nurses in oncology care.

3.2. Nurses Have Changed Documentation Practices

According to some of the interviewed physicians, many nurses are more restrictive regarding what to write in the medical record now when the patients can access their health records online. This could, in some cases, lead to important information being left out of the medical record. A few of the interviewed physicians presented short incidents related to patients or relatives that became very upset when nurses had added a short note about the patient's mental health status - information that was important for the current treatment. In two cases, strong reactions on information found in the PAEHR had led to nurses deleting relevant information from the medical record.

3.3. The Log List Functionality Makes Nurses Being Questioned

In Region Uppsala, patients who log in to their PAEHR can use a function called *log list* to see which clinicians who have accessed their record and when access has been made. This particular function was brought up during some of the interviews as being problematic especially for nurses.

One of the issues brought up stems from the fact that patients often know the name of their primary physician, but not necessarily the name of the nurses connected to the care team around the patient. Nurses often access patients' medical records by request of a physician and they also often do routine checks of blood work or similar. In all these cases there is a chance that the patient will find unfamiliar names in the log list online. Some physicians reported that nurses were worried about this and they also reported on some cases when patients had asked a physician about unfamiliar names found in the log list. In all cases reported during the interviews, however, these kinds of questions from patients could be easily handled.

Another issue brought up in two of the interviews, related to documentation, is that some nurses often leave small snippets of information in the medical record to explain why access was made. These snippets do not add anything that could be seen as medically relevant, and hence these physicians were annoyed by them. The occurrence of these snippets was attributed to the existence of the online log list and interestingly enough the issue was brought up spontaneously by these two physicians before the log list had been brought up by the interviewer.

4. Discussion

The results presented in this paper highlight some effects of PAEHR on the work situation of nurses in oncology care. This is one of the few papers to date that reports on effects of PAEHR for nurses in any discipline and also one of the very first that covers effects several years after the launch of *Journalen* in Sweden. It is important to note that the number of interviewed nurses is relatively small. Even though clear problem areas have been identified, more research is needed that include a larger sample as well as interviewees from different medical disciplines.

It is clear from the results of the recent interview study that there are some clearly identifiable effects related to *patient contact*, *documentation practices* and *the access log list functionality*.

When it comes to effects of PAEHR on the contact with patients, both nurses and physicians experience that PAEHR has resulted in nurses receiving more questions from patients. The increased workload that is described could among other things be explained by the gendered norms related to the nursing profession. Nursing is traditionally coded as a female profession quite low in the hospital hierarchy, and the traditional terms that define nursing are emotional and nurturing work [9]. The contact between nurses and patients is also more often continuous and close, and the expectations for nurses are quite different from that on physicians when it comes to communication [15]. Physicians' meetings with patients, on the other hand, are more often fragmented, and short term. These differences among professions, the lack of research on effects of PAEHRs for nurses, as well as the pronounced effects for nurses shown in this study, highlights the importance of including nurses in these kinds of studies.

However, one should note that this increased workload for nurses when it comes to questions from patients might not be only negative in nature, as other studies have shown that person-centeredness in care has correlations with nurses' improved satisfaction with care and work [16]. More research is needed when it comes to the correlation between job satisfaction and the launch of PAEHR.

The perceived effects on documentation practices could also be a result of different gendered expected behaviours between the professions, looking at nursing as a traditionally female coded profession. Changed documentation practices might be understood in the light of women's more relational way of thinking. This indicates that women, more than men, notice and respond to other people's needs and expectations [17] [18]. In this case, nurses change and respond to the patient's need and expectations to understand PAEHR to a larger extent than physicians do as one way of doing gender [8]. This result could have implications for future nursing education especially when it comes to balancing the need to keep a complete and accurate record and the need to satisfy patients.

The log list function has been discussed in earlier research when it comes to effects of PAEHR on work, and especially perceived risks of threats and violence towards medical staff [6]. The results from the present interview study do not give any indication that nurses perceive any risk of threats or violence, but the existence of the online log list still has an impact on actions taken by nurses. The very nature of the nurse profession - in this case being a part of a team who may work with many doctors, and often not

being known by his/her name by the patient - is at the core of the problem. There is a clearly described insecurity here in relation to documentation, and an obvious need to explain one's presence in the electronic health system. More research is needed into the gendered explanations of this in the sociotechnical system.

Acknowledgements

This research was funded by Forte (Swedish Research Council for Health, Working Life and Welfare) [grant number 2016-07153].

References

- [1] D. Allen, The invisible work of nurses: Hospitals, organisation and healthcare, 2014. doi:10.4324/9781315857794.
- [2] A.M. Urban, Taken for granted: Normalizing nurses' work in hospitals, *Nurs. Inq.* (2014). doi:10.1111/nin.12033.
- [3] D. Golay, An invisible burden: An experience-based approach to nurses' daily work life with healthcare information technology, (2019).
- [4] T. Ålander, and I. Scandurra, Experiences of Healthcare Professionals to the Introduction in Sweden of a Public eHealth Service: Patients' Online Access to their Electronic Health Records, in: *Stud. Health Technol. Inform.*, 2015. doi:10.3233/978-1-61499-564-7-153.
- [5] Å. Cajander, J. Moll, S. Englund, and A. Hansman, Medical records online for patients and effects on the work environment of nurses, in: *Stud. Health Technol. Inform.*, 2018. doi:10.3233/978-1-61499-852-5-271.
- [6] U.S. Åkerstedt, Å. Cajander, J. Moll, and T. Ålander, On threats and violence for staff and patient accessible electronic health records, *Cogent Psychol.* (2018). doi:10.1080/23311908.2018.1518967.
- [7] M.A. Johansen, P.E. Kummervold, T. Sørensen, and P. Zanaboni, Health professionals' experience with patients accessing their electronic health records: Results from an online survey, in: *Stud. Health Technol. Inform.*, 2019. doi:10.3233/SHTI190273.
- [8] C. West, and D.H. Zimmerman, Doing gender, *Gen. Soc.* (1987). doi:10.1177/0891243287001002002.
- [9] S. Lindsay, The care-tech link: An examination of gender, care and technical work in healthcare labour, *Gender, Work Organ.* (2008). doi:10.1111/j.1468-0432.2007.00387.x.
- [10] E.M. Rogers, Diffusion of innovations: Simon and Schuster., Computer (Long. Beach. Calif). (2010). doi:10.1109/2.48.
- [11] C. Grönloh, G. Myreteg, A. Cajander, and H. Rexhepi, "why do they need to check me?" patient participation through ehealth and the doctor-patient relationship: Qualitative study, *J. Med. Internet Res.* (2018). doi:10.2196/jmir.8444.
- [12] C. Grönloh, A. Cajander, and G. Myreteg, The record is our work tool-physicians' framing of a patient portal in Sweden, *J. Med. Internet Res.* (2016). doi:10.2196/jmir.5705.
- [13] Å. Cajander, C. Grönloh, and H. Rexhepi, Worlds Apart-Doctors' Technological Frames and Online Medical Records, in: *INTERACT 2015 Adjun. Proc. 15th IFIP TC. 13 Int. Conf. Human-Computer Interact.* 14-18 Sept. 2015, Bamberg, Ger., 2015: p. 2015357.
- [14] G. Erlingsdottir, C. Lindholm, and T. Ålander, eHealth services, patient empowerment and professional accountability-An empirical study on the changing patient-doctor relationship in the digital world, in: *Int. EIASM Public Sect. Conf.*, 2014: pp. 1-21.
- [15] C. Foss, Gender bias in nursing care? Gender-related differences in patient satisfaction with the quality of nursing care, *Scand. J. Caring Sci.* (2002). doi:10.1046/j.1471-6712.2002.00045.x.
- [16] A. Lehlumante, A. Nilsson, and D. Edvardsson, The influence of a person-centred psychosocial unit climate on satisfaction with care and work, *J. Nurs. Manag.* (2012). doi:10.1111/j.1365-2834.2011.01286.x.
- [17] K. Waerness, Caring as women's work in the welfare state, *Patriarchy a Welf. Soc. Oslo Univ.* (1984) 67-87.
- [18] C. Gilligan, In a different voice: Psychological theory and women's development, Harvard University Press, 1993.

Online Medication Information for Citizens: A Comparison of Demands on eHealth Literacy

Helen MONKMAN^{a,1}, Thomas SCHMIDT^b, and Christian NØHR^b

^a*School of Health Information Science, University of Victoria*

^b*The Maersk Mc-Kinney Moller Institute, University of Southern Denmark*

Abstract. Many people take prescription medications and need information about the risks and benefits associated with taking them. Citizens are increasingly turning to the internet for health information and medication information is no exception. There are a variety of websites that offer Online Medication Information for Citizens (OMIC). This study compared six such websites using the Health Literacy Online (HLO) Checklist as a framework. Additionally, we conducted a detailed analysis of the individual content in each OMIC for three different medications. We identified several strengths and weaknesses of the different websites in terms of how they were designed and written and their appropriateness for users with limited eHealth literacy.

Keywords. Consumer health informatics, medication information, health literacy, eHealth literacy

1. Introduction

Worldwide spending on pharmaceuticals (i.e., prescription medications) reached 1.2 trillion in 2018 and is projected to continue increase in the coming years [1]. Many North Americans use short-term and/or long-term prescription medications and use invariably increases with age [2, 3]. Given the prevalence of prescription medication use and the risks and benefits of taking medications, people need information about them. The most commonly used resources for medications are patient information leaflets, doctors, and pharmacists [4]. However, the internet gaining popularity as a resource for medication information, especially younger people [4].

This study assessed how well Online Medication Information for Citizens (OMIC) was designed for users with limited health or eHealth literacy in terms of both its content and design. Specifically, three prescription medications from six online resources were examined to determine their relative strengths and weaknesses using the Health Literacy Online (HLO) Checklist [5] as a framework with a complementary in-depth examination of each OMIC's individual content [6]. HLO is provides evidence-based guidance to assist in the development of "user-friendly health websites and digital health tools" [6].

¹ Helen Monkman, E-mail: monkman@uvic.ca

2. Methods

We assessed OMIC from six different Websites for three different medications: Asmanex (mometasone inhalation), Crestor (rosuvastatin) and Humalog (insulin lispro) (see Table 1).

Table 1. Websites Examined Offering Online Medication Information for Citizens.

Website	Type of Resource	Medication Information Data Source
Rexall.ca	Canadian pharmacy	MediResource Inc.
Walgreens.com	American pharmacy	Clinical Drug Information LLC
Min.Medicin.dk	National Danish medication information resource	Danish Medicines Information A / S
WebMD.com	Health information resource	First Databank, Inc.
MayoClinic.org	Not-for-profit health information resource	IBM Watson Micromedex
Drugs.com	Independent medication information resource	IBM Watson Micromedex, Cerner Multum™, Wolters Kluwer™ and others

We compared each website against the four applicable categories of the Health Literacy Online Strategies (HLO) Checklist: *Write Actionable Content*, *Display Content Clearly on the Page*, *Organize Content and Simplify Navigation*, and *Engage Users* [5]. The category *Test Your Site with Users with Limited Literacy Skills* was excluded because this was an inspection study.

We used word count and Simple Measure of Goobledygook (SMOG) scores [7] as indices of HLO Checklist's guidance to write in plain language in the Write Actionable Content Category. SMOG [7] scores are derived using a calculation of the number polysyllabic words in a sample sentences to estimate of how many years of education are required to understand the text. To this end, Danish OMIC from min.medicin.dk, was translated to English using Google Chrome and each OMIC was converted to a simple text document (either through print functionality or copying and pasting). We then determined each OMIC's SMOG score using an online readability calculator [8]. Notably, we excluded the information on the "professional" tab of Drugs.com, as it was not for the target audience of citizens. We calculated the average number of words and average SMOG scores for each website.

Additionally, to assess the content in greater detail, we used a previously developed set of criteria for content utility which outlines 11 helpful topics for medication information: benefits, contraindications, directions, precautions, adverse effects, storage, missed dose, allergic reaction, drug interaction, overdose, and general information [6]. We averaged the number of content criteria satisfied amongst the three OMIC per website. Further, we examined each OMIC's reported side effects to determine whether there was conflicting medication information [9].

3. Results

3.1. Write Actionable Content

Min.medicin.dk (1,296 words) and Walgreens (1,501 words) used the fewest words on average for their OMIC. Drugs.com was the lengthiest, with an average of 4,507 words. Walgreens had the most readable content, with the lowest average SMOG score of 8.47. The remaining five websites' average SMOG scores ranged from 11.69 for

WebMD to 14.51 for Drugs.com. As alluded to in the methods, Drugs.com also contained information for professionals in the form of a tab and as additional content in its side effects tabs. This has the potential to create confusion for users and unnecessarily overwhelm them. In contrast, Min.medicin.dk offers an alternative site specifically for professionals (pro.medicin.dk). Drugs.com and Rexall had many bullets and lists, whereas others had fewer (e.g., WebMD, MayoClinic) or almost none (i.e., Walgreens).

3.1.1. Content Utility and Comparison of Potential Side Effects

Min.medicin.dk satisfied the fewest content criteria (73%), followed by Rexall (82%), Walgreens (88%), Drugs.com (88%), and the MayoClinic (91%). WebMD (97%) satisfied the most content criteria.

There was considerable variation between which possible side effects were reported, how many in total, as well as how they categorized the potential side effects. Specifically, Min.medicin.dk reported the fewest possible side effects amongst all of the websites, whereas the MayoClinic and Drugs.com reported the most. There were also discrepancies with the language used to describe or categorize possible side effects: some websites focused on frequency of occurrence (e.g., more common, less common, rare, may occur, unlikely) and others focused on severity (e.g., very bad, serious) or both frequency and severity. Arguably, Min.medicin.dk utilized the clearest approach to describe the frequency of side effects by using numbers. For example, they reported maximums of “10 out of 100 people get the side effect” (common), or “1 out of 1 people” (uncommon) and “1 out of 1,000 people (rare) as well as unknown frequencies of possible side effects. In terms of text presentation, Walgreens was the least structured and most difficult to read and extract information due to its format. Specifically, the information was all presented in big blocks of text with capitalized letters to attract attention or serve as headings. Additionally, Drugs.com listed additional possible side effects and frequencies for health care professionals that were categories by medical specialty (e.g., gastrointestinal, nervous system, respiratory) and labelled more common (10% or more), common (1% to 10%) or uncommon (0.1% to 1%). Some sites listed side effects that do not require medical attention, but which may still cause consumer concern.

There was no consensus on which possible side effects were reported or how they were reported. For an illustrated example, to compare whether or not a possible side effect was reported by each site and how it was labeled is provided in Table 2.

Table 2. A Comparison Sample of Possible Side Effects for Reported by Each Website for Asmanex

Website	Fever	Cough	Nausea	Headache
Rexall.ca	*	✓	✓	✓
Walgreens.com	✓	✓		✓
Min.Medicin.dk				✓
WebMD.com	*	✓		
MayoClinic.org	✓	✓	✓	✓
Drugs.com	✓	✓		✓

Legend: ✓ = Possible side effect reported. * = Reported as a sign of infection, not as a possible side effect.

3.2. Display Content Clearly on the Page

Some sites displayed their content more clearly than others. Specifically, Walgreens and Min.medicin.dk were both spacious and visually appealing. They achieved this by

using collapsible categories. The selected categories varied greatly amongst sites. Walgreens listed just five selection options (Instructions, Side effects, Patient reviews, Pricing, and PDF download). At the other end of the spectrum, min.medicin.dk lists 16 selection options such as active compounds, packaging, precautions related to breast feeding, pregnancy, and donating blood. Drugs.com contained several advertisements which not only gave it a busier appearance but could potentially distract users from the main content. Similarly, WebMD had many links, which could encourage users to navigate away from the page before reading its entirety. Further, on almost every tab on WebMD, users had to click the “read more” button to reveal the remaining content. WebMD also had an excessive amount of links and not enough bullets or short lists. With the exception of Rexall, nearly all of the websites provided some helpful images of pictures of the medications. However, Min.medicin.dk leveraged multimedia further for medication instructions. MayoClinic and Min.medicin.dk provided buttons and printer friendly versions of the OMIC. Walgreens allows users to download PDFs of the OMIC and Rexall can be printed using the browser. However, WebMD and Drugs.com did not offer printer friendly options.

3.3. Organize Content and Simplify Navigation

Search functionality varied amongst the six resources. Min.medicin.dk, Drugs.com, Walgreens and WebMD demonstrated strong search functionality by supporting users with auto-complete and offering suggestions for misspellings. However, Rexall and WebMD returned no results for misspelt medications. Moreover, Drugs.com and WebMD search results were displayed below advertisements and therefore disobeyed the HLO Checklist [5] guidance to display search results clearly. Despite having autocomplete functionality, the medications were not suggested by the MayoClinic. Further, the MayoClinic’s search results only named the generic name of the medication and not the brand name, which could lead users to falsely conclude their search was unsuccessful. We tested the responsiveness of all six websites and found to be present using the developer functions of Google Chrome.

3.4. Engage Users

Only Rexall and Drugs.com demonstrated strong sharing capabilities by providing options on each OMIC page to share the information on social media (e.g., Facebook, Twitter). Min.medicin.dk did not support sharing to social media, but it did provide an option to email the link, as did Drugs.com. Walgreens (sourced from PatientsLikeMe) and Drugs.com also included patient reviews of the medications on their sites.

4. Discussion and Conclusion

There were several strengths and weaknesses of the six websites in this investigation. Overall, the strongest site in terms of placing the fewest demands on users (i.e., abided by the guidance of the HLO Checklist) was Walgreens and its content was the most readable and easy to navigate. However, despite performing well on the metrics, the information itself poorly formatted. Specifically, Walgreens content was frequently dense walls of text. Thus, it is important to note that despite satisfying many criteria, there is still substantial room for improvement of its content (e.g., use bullets and short

lists). Moreover, are some HLO Checklist criteria more important (i.e., impact users more) than others? Should they be weighted?

Alarming, there were discrepancies between how the websites reported possible side effects. Specifically, some reported possible side effects, while other websites did not report the same ones (see Table 2). Conflicting medication information has been shown to negatively impact adherence [9] and therefore there are likely consequences to these OMIC side effect discrepancies. Further, how websites categorized varied by emphasizing frequency of occurrence or severity of experiencing the side effects, concepts often confused by citizens [10]. Moreover, how frequency of occurrence using either a vague label (e.g., common) or specific value (e.g., 1 in 100 people) could have an effect on citizens' interpretation and might also impact adherence.

There were some limitations to this study. We only examined three medications from the six websites. Further, we excluded pharmaceutical manufacturer information. The Danish OMIC was translated using Google Chrome which is an imperfect proxy. Additionally, websites that did not provide a print option (i.e., Drugs.com, WebMD) were copied and pasted and excluded the redundant content that around the tab areas. Finally, we neither tested the accessibility nor the mobile versions of the websites.

There are a variety of OMIC options available to citizens and each website in this study had its strengths and weaknesses. Based on the current evaluation, we recommend that these websites make an effort to improve the consistency of information they are providing to citizens, to limit conflicting medication information. Further, findings from this study could be incorporated to improve the OMIC and its adherence to the HLO guidelines for simplicity of navigation and content itself and also its structure are warranted. Future research opportunities including comparing the influence of health care system structure, culture, and prescription practices internationally (e.g., North America vs. Europe), as how citizens select amongst OMIC alternatives, and finally how these resources might be designed to address the needs of people with varying eHealth literacy skills.

References

- [1] IQVIA Institute for Human Data Science. The Global Use of Medicine in 2019 and Outlook to 2023: Forecasts and Areas to Watch, Institute Report, January 29, 2019.
- [2] M. Rotermann, C. Sanmartin, D. Hennessy, M. Arthur. Prescription medication use by Canadians aged 6 to 79. *Health Matters. Health Rep.* **25**(6) June 2014, 3-9. Statistics Canada, Catalogue no. 82-003-X.
- [3] IQVIA Institute for Human Data Science. Medicines Use and Spending in the U.S.: A Review of 2016 and Outlook to 2021, Institute Report, May 4, 2017.
- [4] U. Närhi, Sources of medicine information and their reliability evaluated by medicine users, *Pharm World Sci.* **29** (2007) 688–694. doi:[10.1007/s11096-007-9131-1](https://doi.org/10.1007/s11096-007-9131-1).
- [5] Office of Disease Prevention and Health Promotion. Checklist: Health Literacy Online. Retrieved, September 22, 2019 from <https://health.gov/healthliteracyonline/checklist/>
- [6] H. Monkman, and A.W. Kushniruk, Consumer Medication Information: Similarities and Differences Between Three Canadian Pharmacies, *Stud Health Technol Inform.* **234** (2017) 238–242.
- [7] G. H. McLaughlin. SMOG grading-a new readability formula. *J Reading* **12**(8), (1969), 639-646.
- [8] Online-Utility.org: Utilities for Online Operating System. Retrieved September 21, 2019 from https://www.online-utility.org/english/readability_test_and_improve.jsp
- [9] D.M. Carpenter, R.F. DeVellis, E.B. Fisher, B.M. DeVellis, S.L. Hogan, and J.M. Jordan, The effect of conflicting medication information and physician support on medication adherence for chronically ill patients, *Patient Educ Couns.* **81** (2010) 169–176. doi:[10.1016/j.pec.2009.11.006](https://doi.org/10.1016/j.pec.2009.11.006).
- [10] V. Tong, D.K. Raynor, S.J. Blalock, and P. Aslani, Exploring consumer opinions on the presentation of side-effects information in Australian Consumer Medicine Information leaflets, *Health Expect.* **19** (2016) 543–556. doi:[10.1111/hex.12215](https://doi.org/10.1111/hex.12215).

Personalized Communication - People with Hearing Loss' Experiences of Using a National Patient Portal

Maria HÄGGLUND^{a, b 1}, Emma HELSING^a, and Isabella SCANDURRA^c

^a*Department of Women's and Children's Health, Uppsala university, Uppsala, Sweden*

^b*Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, USA*

^c*Informatics, School of Business, Örebro University, Örebro, Sweden*

Abstract. Patient portals are used as a means to facilitate communication, performing administrative tasks, or accessing one's health record. A qualitative study was therefore performed to understand how people with hearing loss experience using the Swedish national patient portal 1177.se and the patient accessible electronic health record *Journalen*. Data was collected through focus groups and individual interviews. Ten persons with varying degree of hearing loss, age and years of using a hearing aid participated. The data was analyzed through thematic content analysis and three themes emerged; access to written information, patient-created information and personalized communication. Although the study is limited in size, we believe that important lessons can be learned regarding the needs for personalized communication, including access to written information.

Keywords. Patient portals, Electronic health records, Hearing loss

1. Introduction

Patient portals are used to facilitate communication between patients and healthcare professionals, as well as for performing administrative tasks, such as appointment bookings and prescription renewals. Patients are also increasingly provided with access to their electronic health records (EHRs), sometimes referred to as patient accessible EHRs (PAEHRs) [1][2] or OpenNotes [3], through portals. Most of these systems are designed to meet the needs of the general population, and few studies are available on how patients with special needs experience the use of such systems. In Sweden, a national patient portal is used that connects to all EHR systems used in the different regions [2], through a national health information exchange platform [4]. By logging in using a national e-ID one can access a number of administrative services as well as the PAEHR *Journalen*. In July 2019, 61% (6.3 million/10.3 million) of the Swedish population had logged in to the portal, and 29% (3 million/10.3 million) have used the PAEHR *Journalen* at least once [5]. Several studies have been performed exploring patients' and healthcare professionals' views on *Journalen* [6][7], but to our knowledge, no studies have focused on how persons with hearing loss use and experience the patient portal 1177.se and *Journalen*. Since this group of users have additional needs regarding

¹ Corresponding Author, Maria Hägglund, Department of Women's and Children's Health, Uppsala University, Dag Hammarskjölds väg 14B, 752 37 Uppsala, Sweden. Email: maria.haggglund@kbbh.uu.se.

communication and access to information, we aim to explore their experiences of using the Swedish national patient portal 1177.se and the PAEHR *Journalen*, as well as their suggestions for future improvements.

2. Methods

Hearing aid users who had used 1177.se and *Journalen* were recruited through online advertisements spread through social media and patient organizations. The participants (n=10) were recruited from different parts of Sweden, and given the option to join a focus group interview or to participate in an individual interview, in person or over the phone. Two focus groups were conducted; in Uppsala and Stockholm. Each focus group lasted 90 minutes. Three individual interviews were conducted by telephone and lasted approximately 40 minutes. A validating workshop was conducted using raw data from focus groups and interview, confirming the results that emerged and is therefore not reported separately. The participants represented a variation of age, gender, severity of hearing loss and years of using hearing aids (Table 1).

Table 1. Demographics of the participants (n=10). U = Uppsala FG, S = Stockholm FG, I = Interview, V = Validating workshop

Participant ID	Gender	Age	Years using hearing aids
U1	Male	47	10
U2	Male	76	5
S1	Male	51	44
S2	Female	40	12
S3	Female	47	42
I1	Male	69	6
I2	Male	82	16
I3	Female	75	20
V1	Female	75	15
V2	Female	74	1

Data were analyzed using a thematic content analysis approach, providing a rich and thorough, interpretation of the data [9]. The analysis was done manually and in iterative dialogue between the authors. The Swedish Ethical Review Authority granted ethical approval for the project PACESS (Dnr: 2017/1028-31).

3. Results

Both current experiences of using 1177.se/*Journalen* and participants’ suggestions for future improvements were discussed. Three themes were identified; access to written information, patient created information and personalized communication.

3.1. Access to written information

Having access to written information in general and the notes from the electronic health record, in particular, was described as very important by the participants. Depending on the gravity of the participants hearing loss, it required a lot of energy and focus to listen to what clinicians said during an appointment. Some of the participants described that they focused so hard on hearing what the person said that they couldn’t process and

remember much of the information. Being able to go back and review notes then became even more important.

Access to the record as a safety tool. A few of the participants described that having access to record notes was also a patient safety issue, as they could see when the information was missing or wrong.

"When I saw that things were missing in my health record I contacted the clinic and asked why." (U1)

Improved understanding with access to record notes. Reading the health record online gives the patient a summary of what was said during the meeting. Having the option to read their health record online was described as a means to improve communication with healthcare.

"I get a summary of what we have been talking about during the care visit. Things that I have misunderstood during the actual meeting, I can go in and read about in my health record to understand what the healthcare professional meant." (U1)

Specific information for hearing aid users. Several participants reasoned that they would like the hearing clinic to post more specific information for hearing aid users on 1177.se, e.g. related to specific services that hearing aid users may benefit from.

"New hearing aid users are extremely dependent on their audiologist and the information provided by the audiologist. More information about this should be available at 1177.se." (S1, S2, S3)

A few of the participants would also like self-management films on 1177.se on e.g. how to care for the hearing aid.

Diagnostic examination results. All the participants had ideas for what kind of diagnostic examination results they would like to see on 1177.se; audiograms, X-rays and ECG (this type of information is currently not available). Some of the participants received paper copies of their audiogram when they visited the audiologist others received the audiogram as a pdf file to their email. Several participants would rather have it accessible through 1177.se

"I received my audiogram on paper, and after the visit, I took a picture of it with my phone, so I could have it accessible, but I would rather have it on 1177.se." (S2)

3.2. Patient created information

The participants did not only want access to more and better information, they also wanted to contribute with their own information to healthcare through 1177.se. Three different types of patient created information was suggested.

Add attachments to messages. In one of the focus groups, the participants wanted to send attachments with messages through 1177.se, for example, sending medical certificates in attachments.

"I had contact with my counselor through e-mail and she asked if we could take the errand through 1177.se instead but that was a problem because I was not able to send attachments with my medical certificate through 1177.se" (U1)

Add information about hearing and medical aids. A list or a profile where patients could add information about what hearing and medical aids they use was also suggested. Having a list or a profile makes it easier for other healthcare providers or other audiologists to know what the patient already use.

"I would be good if there were some kind of list on 1177.se of what hearing aids and medicals I use so that I could go in and add things and perhaps write instructions how my different aids work and how they are set." (U1, U2)

Commenting/adding to their own health record. At the moment, the only functionality for most users of 1177.se/Journalen is the ability to read their health records. In the Uppsala Region, it is, however, possible for patients to make comments in the record that healthcare professionals can read. A few of the participants thought that this function would help them reflect and remember things about their health and that it could also save time during the actual health meeting.

"When I read my health record entries, I get ideas and remember things about my health from several years ago and this I would want to add to my health record. It would also save time for both myself and the healthcare provider. /.../ it would also help me bring this up in the next meeting" (U1)

3.3. Personalized Communication

Communication with healthcare and the importance of the national patient portal for users with hearing loss in this communication was an important topic.

Impact of hearing loss on communication with healthcare. Several participants reasoned about how their lack of hearing affected their communication with healthcare. Many did not use the telephone when contacting healthcare since their hearing loss increases the risk for misunderstandings, and were happy that 1177.se provides means for written communication instead.

"I never use the telephone when contacting healthcare because I have a really hard time hearing what they say, I use 1177.se so I don't miss important things." (S2)

Yet, several of the participants expressed frustration that despite their hearing loss, healthcare would insist on using the telephone when initiating contact.

"For me, it doesn't work at all. My health record from the hearing aid clinic states that I am deaf and cannot receive phone calls. SMS or messages via my 1177.se works but despite this, I get calls from healthcare professionals." (U1)

Participants with less severe hearing loss used the telephone to communicate with healthcare, and it was clear that different individuals had very different needs.

Lack of dialogue through 1177.se. Users of the patient portal 1177.se can send messages to their healthcare providers. Several of the participants reasoned that sending and receiving messages worked well through 1177.se but after receiving an answer the dialogue ended, and they could not send anything back.

"I think it's good that you get answers, but there is no dialogue or feedback. It's difficult to have a discussion if there's something you want to explain or bounce ideas. It works when you want to book lab tests or renew prescriptions." (S2)

Communication and equality. Participants described different ways of communicating with their audiologist/hearing clinic. Some had direct email contact with their audiologist and appreciated this but were aware that this was a special case.

"The audiologist does not give out their e-mail address to whoever. /.../ But I have contact with my audiologist via e-mail." (S2)

Alternative online communication methods. Participants wanted to be able to send messages directly to the audiologist through 1177.se, but they also proposed a chat function with receptionist/nurse or using a videophone.

"Videophone, sign language chat should be offered." (S1, S2, S3)

Handling appointments. Currently, patients receive appointments from many healthcare professionals in a variety of formats (oral, paper and in some cases electronic). Many of the participants would prefer booking appointments online.

"Appointment booking to my audiologist would be better if it was digital." (I3)

Others reasoned that there should be an easier booking system in 1177.se and that healthcare professionals should be able to send appointment bookings digitally.

“Easier appointment booking system on 1177.se like a guide that would help you if you not sure how to book an appointment.” (S1, S2, S3)

Ensure awareness of a person's hearing loss among healthcare professionals. Many of the participants were concerned that healthcare professionals did not always know that they had a hearing loss. Information should be clearly visible in the patient's health record and healthcare professionals' portal interface, perhaps in the format of a patient profile, to both reduce the risk for communication problems and save time during visits, for now, hearing-related health issues.

“Other healthcare professionals should know I have hearing loss.” (S1, S2, S3)

4. Discussion and conclusion

Although a fairly small group of people participated, the in-depth interviews provided a rich material that clearly indicates some of the challenges experienced by people with hearing loss in their communication with healthcare, and highlights benefits that digital tools can and do provide for these users. Further flexibility and adaption to individuals' needs and preferences are still needed to fully support people with hearing loss. Improved tools for written communication and access to clinical documentation can provide such support, and would likely be beneficial to other users.

Acknowledgement FORTE – the Swedish Research Council for Health, Working Life and Welfare funds the research project “PACESS” (2016-00623). This research was also supported by Uppsala MedTech Science & Innovation (www.medtech.uu.se), a joint strategic initiative between Uppsala University and Uppsala University Hospital.

References

- [1] D. Wiljer et al., “Patient accessible electronic health records: Exploring recommendations for successful implementation strategies,” *J. Med. Internet Res.*, vol. 10, no. 4, 2008.
- [2] M. Hägglund and I. Scandurra, “Patients' online access to Electronic Health Records – current status and experiences from the implementation in Sweden,” *Full Pap. Accept. Medinfo 2017*, 2017.
- [3] J. Walker et al., “OpenNotes After 7 Years: Patient Experiences With Ongoing Access to Their Clinicians' Outpatient Visit Notes,” *J Med Internet Res*, vol. 21, no. 5, p. e13876, 2019.
- [4] N. Davoody, S. Koch, I. Krakau, and M. Hägglund, “Accessing and sharing health information for post-discharge stroke care through a national health information exchange platform - a case study,” *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 95, 2019.
- [5] Inera, “Inera Statistics.” [Online]. Available: <https://www.inera.se/aktuellt/statistik/>.
- [6] J. Moll et al., “Patients' Experiences of Accessing Their Electronic Health Records: National Patient Survey in Sweden,” *J Med Internet Res*, vol. 20, no. 11, p. e278, 2018.
- [7] C. Grunloh, A. Cajander, and G. Myreteg, “‘The Record is Our Work Tool!’-Physicians' Framing of a Patient Portal in Sweden,” *J. Med. Internet Res.*, vol. 18, no. 6, p. e167, Jun. 2016.
- [8] L. Petersson and G. Erlingsdottir, “Open Notes in Swedish Psychiatric Care (Part 1): Survey Among Psychiatric Care Professionals,” *JMIR Ment. Heal.*, vol. 5, no. 1, p. e11, Feb. 2018.
- [9] U. H. Graneheim and B. Lundman, “Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness,” *Nurse Educ. Today*, vol. 24, pp. 105–112, 2004.

Process Approach for Managing Health Information System-Induced Medication Errors

Maryati MOHD. YUSOF^{a,1}, Toshihiro TAKEDA^b

Naoki MIHARA^c and Yasushi MATSUMURA^b

^a*Faculty of Info. Sc. and Technology, Universiti Kebangsaan Malaysia, MALAYSIA*

^b*Dept of Medical Informatics, Osaka University Graduate School of Medicine, JAPAN*

^c*Department of Medical Informatics, National Cancer Center Hospital, JAPAN*

Abstract. Health information systems (HIS) and clinical workflows generate medication errors that affect the quality of patient care. The rigorous evaluation of the medication process's error risk, control, and impact on clinical practice enable the understanding of latent and active factors that contribute to HIS-induced errors. This paper reports the preliminary findings of an evaluation case study of a 1000-bed Japanese secondary care teaching hospital using observation, interview, and document analysis methods. Findings were analysed from a process perspective by adopting a recently introduced framework known as Human, Organisation, Process, and Technology-fit. Process factors influencing risk in medication errors include template- and calendar-based systems, intuitive design, barcode check, ease of use, alert, policy, systematic task organisation, and safety culture. Approaches for managing medication errors also exert an important role on error reduction and clinical workflow.

Keywords. Medication error, health information systems, evaluation, workflow, lean, case study

1. Introduction

Despite the potential of health information systems (HIS) in reducing medication errors, HIS have been associated with new types of medication errors [1]. This paradox is primarily attributed to HIS misfit with the work patterns and settings of health care, resulting in the inefficient use and unintended impact of HIS in error reduction. The prevalence of HIS adoption and safe application remains at its infancy due to technical and organisational challenges [2, 3]. Further research is required on work organisation problems, cooperative work problem [2] and identifying safe practices for managing information technology (IT) transition. A close look at clinical process management can demonstrate the overall process, its specific steps in identifying the control mechanism and error risk and its impact pertinent to HIS-induced errors. This paper reports the preliminary findings of a case study on the management of HIS-induced medication errors from a process perspective in clinical practice.

¹ Corresponding Author. Maryati.Yusof@ukm.edu.my

2. Theoretical background

The proposed Human, Organisation, Process and Technology-fit (HOPT-fit) evaluation framework was developed after critically appraising the literature [4]. Previous error models were also used to categorise evaluation factors, dimensions and measures. As an organisational element, process is featured as one of the factors and is represented as the dashed line that links process and organisation in Figure 1. Process plays a central role in error failure and management because errors are commonly triggered during the execution of a process. The developed framework proposed three dimensions of process: the clinical stages, business process management (BPM) life cycle and lean methods. This study focused on medication errors; thus, it examined the stages of medication and its compliance with the five rights: right drug, dose, route, time and patient (5R). Process management can be assessed in accordance with various stages of BPM. Meanwhile, process quality and safety can be examined using various lean tools and methods.

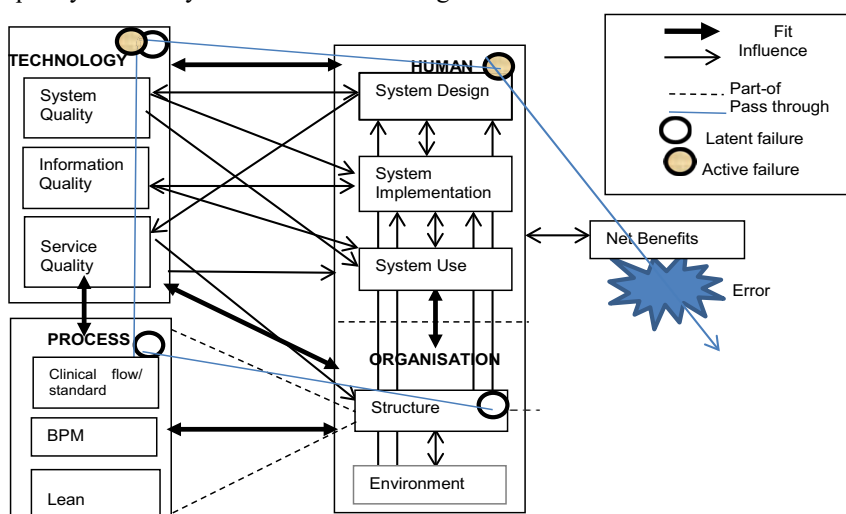


Figure 1. HOPT-fit framework (adapted from [4]).

3. Method

A summative case study evaluation was conducted in actual clinical settings to explore the context and nature of the event to be investigated. Data were gathered using interview, observation and document/artefact analysis. We studied a number of HIS in a 1000-bed Japanese secondary care teaching hospital (SCTH). Face-to-face and written interviews were audio- and manually recorded and then transcribed in English and Japanese by translators. Using a purposeful snowball sampling method, 13 informants among clinicians and management committee were interviewed. Four techniques were used to analyse the results: coding, analytic memos, contextual analysis and narrative analysis. To understand medication error management holistically, data were analysed using process approaches [5, 6]: 1) Major medication processes in an end-to-end clinical process were identified and structured into five sub processes, from prescribing to monitoring. 2) A process step table was constructed; it includes process information flows and owners. 3) Process error risks were summarised, and their sources (input,

process or output flaws) and subsequent impact were identified. 4) Activities that serve as preventive, detective or corrective controls were determined.

4. Results

Medication use (MU) for the SCTH's inpatients was documented using a process step table that featured the medication process, control, flaw, and impact. MU steps were identified based on the MU process for hospital and long-term cases [7]. The SCTH additional processes is nearly double the number of those of MU. The selected steps, error control, flaw and impact are provided in Table 1. The medication process for the STCH inpatients involved several HIS, and a safety measure is implemented on every possible point of care. The doctor *prescribes* medicine in the consultation room using a computerised provider order entry (CPOE). The medication order is sent to the pharmacy department for drug preparation. The doctor gives order instructions to the nurse through a calendar, template-based instruction system (InstS). The nurse receives the instruction. The doctor checks whether the nurse has seen the instruction (indicated by a changed coloured display). The nurse sees the new doctor's instructions that are highlighted in different colour. When the nurse checks it, he/she can post questions to the doctor or clicks the 'received button' to confirm the receipt of the doctor's instruction. The pharmacist checks the order data on Pharmacy IS (PhIS), downloads the data to PhIS, prepares the drug envelope and sends the envelope and drug to the nurse station for dispensing the medication to the address entered in a mechanical cart. The nurse delivers the envelope to the ward for administration. STCH policy requires the preparation of a one-dose package to avoid confusion.

Changes in the doctor's order are designed with error control but are still prone to errors. Error risk may occur when the doctor changes the order but the nurses are unaware of it because they do not always check the system. Therefore, the system should ensure that nurses are informed about changes. If the doctor changes instructions before a drug is mixed, then the nurse will print the mixing sheet with a new bar code. Otherwise, if the nurses check the old mixing sheet bar code against the new bar code update in the system, then an error alert is prompted due to bar code mismatch. Similarly, if a change is made after an order is sent to the pharmacy, then the pharmacy does not know about the change. The drug may already be delivered but administration should be stopped. The instruction system is updated with the change and the medicine should be returned to the pharmacy. The new mixing sheet is printed out. The information in the pharmacy system is only updated if drug quantity increases.

5. Discussion

The current study adopted the core processes and adapted them to MU processes that are unique to the SCTH setting. The process model is beneficial for modelling the overall process, ensuring the inclusion of basic processes and arranging them in a structured and systematic manner. Breaking down the process to the simplest task unit enabled the discovery of process variance, barrier, error risk, and fit with HIS that are otherwise overlooked in the SOP. The process of determining step breakdown and classifying it as either a step or control, a single or a sub step is challenging. Optional activities are

explicitly modelled due to their intuitive manner and positive influence on ‘the structural complexity and the understanding of the process flow’ [8].

Table 1. Selected medication process, control, risks and impact

Process steps [resource(s)]	Control and impact	Error risk and impact
Prescribing		
P1.1 Make clinical decisions	Calendar-based interface, graphical symbol and colour - AL	Overlapping medication - VACT
P1.2 Select drug		
P1.3 Determine drug regimen	Automated dosing and calculation – ACT; Alert - A	Alert override - ART Excessive alert - R
<i>P1.6 Stamp prescription</i>	Identity authentication - V	
<i>P1.7 Submit order</i>	Daily-based drug - AT	
<i>P1.8 [O] Change order</i>	Must cancel previous order - A	Incomplete changes - VAT
<i>P1.9[O] Submit order change</i>		
Instructing		
I2.1 Send instruction	Template, calendar-based interface-L; Timing-based drug - LT	
<i>I2.7 Check if order is received</i>	Automated colour-coded change - LAT	
<i>I2.8 [O] Send change order instruction</i>	Prevent instruction overlap – VAT; Change message – ALT; View nurse check - AT Highlights in MD interface - LT	Nurse does not check changes – VACT; Nurse forgets to exclude stopped medication or include added medication - VACT
<i>I2.8 [O] Check seen changes</i>	Disappeared highlights	
Transcribing		
T3.1 Check order data	Correctness check: IF before Ph. check → THEN OK; IF after Ph. check → THEN error - VAT Automated colour-code change - L	Change is made after order is sent to PhIS and drug is prepared - VACT
<i>T3.2 Download data to PhIS</i>		
Dispensing		
D4.1 Enter and screen data	SOP - VACT	
<i>D4.2 Prepare drug envelope</i>	Completeness of patient and drug information - VACT	
D4.3 Prepare, mix and compound drug	5R, patient-drug bar code match: IF before drug mix → THEN OK IF after drug mix → THEN error message – VACT Prepare one-dose package - AT Two independent nurse checks - VA	Error can occur if the system is not checked before injection drug mixing - AT
D4.4 Dispense to nurse station	Dispensing and cart machine - AT	
<i>D4.5 Dispense to ward</i>		
Administering		
A5.3 Nurse verifies order	5R, patient-drug barcode match: IF match, THEN OK; IF mismatch, THEN prompts error alert- VACT	Nurse does not check the system or use the barcode before administering - VACT
A5.4 Administer drug	Colour code for taken, delivered and undelivered drugs -VACT	Nurse forgets to take out stopped drug or add new drug - VACT

Notes: Italicised step = unique to SCTH Ph. = pharmacy; 5R = 5 rights; Data quality measures: validity (V), accuracy (A), timeliness (T), completeness (C), relevancy (R), legibility (L)

STCH has established various controls to ensure patient safety including intuitive design using template, calendar, and automated functions. However, several controls are

also error-prone. Subsequent flaws that penetrate through the aligned control mechanisms from all four HOPT factors will result in an error incident. By contrast, several error risks have no control mechanism. Continuous monitoring and evaluation through lessons learned can improve HIS development and use, fitting it with the clinical process to prevent recurring errors. In addition, safety culture and Japanese Kaizen approach, which focuses on continuous incremental improvement, has been proven to yield significant results over time in STCH.

Although the study was conducted at specific, clinical settings of one hospital over a short duration, the data were collected rigorously and systematically, yielding to in-depth findings that are also applicable to other clinical contexts. Additional feedback was obtained from the participants to clarify and gain more information.

6. Conclusions

Every process step is prone to errors, but several steps have a control mechanism to prevent or minimize error. Human intervention is still required in tandem with system intervention to avoid errors. The case study findings also demonstrate the feasibility of the process approach to address process as one of the essential components of IS, and the fit between the HOPT factors. This evaluation approach is potentially useful to researchers and practitioners for conducting rigorous evaluation studies. Lessons can be learned from this study's findings and recommendations for improving HIS development and process management can be identified to guide the future development of HIS, increasing its effectiveness and improving patient safety.

Acknowledgement. We thank the Sumitomo Foundation Grant, UKM (DIP-2016-033) and Ministry of Higher Learning Malaysia (FRGS/1/2018/ICT04/UKM/02/5).

References

- [1] NQF (National Quality Framework), *Identification and Prioritization of HIT Patient Safety Measures*, 2016.
- [2] NAS (National Academy of Science), *Health IT and Patient Safety: Building Safer Systems for Better Care Committee*, National Academy of Science: Washington D.C. 2012.
- [3] D.F. Sittig, A.Wright, E. Coiera, F.Magrabi, R. Ratwani, D.W. Bates, H. Singh, Current challenges in health information technology-related patient safety, *Health Informatics Journal*, 2018.
- [4] M. Yusof, *A Socio-Technical and Lean Approach Towards a Framework for Health Information Systems-Induced Error*, in Improving Usability, Safety and Patient Outcomes with Health Information Technology, ed. F. Lau, J.A. Bartle-Clar, G. Bliss, E.M. Borycki, K.L. Courtney, A.M-H. Kuo, A. Kushniruk, H. Monkman, A.V. Roudsari, 2019. 508-512.
- [5] M. Dumas, M. La Rosa, J. Mendling, H. Reijers, *Fundamentals of Business Process Management*, Springer-Verlag, Berlin Heidelberg, 2018.
- [6] A.M. Chircu, J.L. Gogan, S.R. Boss, R. Baxter, Medication errors, handoff processes and information quality: A community hospital case study, *Business Process Management Journal*, **19** (2013), 201- 216.
- [7] IOM (Institute of Medicine), *Preventing Medication Errors*, ed. J.P. Aspden, J.L.B. Wolcott, L.R. Cronenwet, 2007.
- [8] C. Natschläger, V. Geist, F. Kossak, B. Freudenthaler, *Optional Activities in Process Flows*. Proceedings of Enterprise Modelling and Information Systems Architecture, Vienna, 2012.

Qualitative Evaluations of mHealth Interventions: Current Gaps and Future Directions

Meghan BRADWAY^{1a,b}, and Kari LEIBOWITZ^c, Kathleen A. GARRISON^d, Lauren HOWE^c, Eirik ÅRSAND^{a,f}

^aNorwegian Centre for E-Health Research, University Hospital of North Norway

^bDepartment of Clinical Medicine, UiT- The Arctic University of Norway, Norway

^cDepartment of Psychology, Stanford University, USA

^dDepartment of Psychiatry, Yale School of Medicine, USA

^eDepartment of Business Administration, University of Zurich, Switzerland

^fDepartment of Computer Science, UiT- The Arctic University of Norway, Norway

Abstract. Psycho-social factors are often addressed in behavioral health studies. While the purpose of many mHealth interventions is to facilitate behavior change, the focus is more prominently on the functionality and usability of the technology and less on the psycho-social factors that contribute to behavior change. Here we aim to identify the extent to which mHealth interventions for patient self-management address psychological factors. By understanding users' motivations, facilitators, and mindsets, we can better tailor mHealth interventions to promote behavior change.

Keywords. Self-management, apps, behavior change, psycho-social factors

1. Introduction

Mobile health (mHealth) technologies (e.g., smartphone apps or wearables), affect patients' self-management (SM), clinical care, and health research. Especially for those with chronic health conditions like diabetes, mHealth enables patients to gather relevant data such as information about blood glucose, diet, and physical activity to better understand their health and make decisions about diabetes SM. With this knowledge at their fingertips, patients are now encouraged to participate in their care by sharing mHealth data with their healthcare providers (HCPs). As patients do this, HCPs will need to adjust their approach to patient care and guidance, and health researchers need to understand how mHealth technologies impact the ways patients and providers work together.

The purpose of most health and mHealth interventions for lifestyle-related health issues, e.g. diabetes, is to facilitate health management and, if necessary, behavior change. Research on mHealth interventions has focused on user experiences, with some pre-post measures of health behavior change, e.g. frequency of blood glucose measurements. However, less attention is placed on users' environments, motivations, or interactions with others [1]. Both internal (e.g., self-efficacy, sense of control, mindsets about health) and external factors (e.g., social

¹ Corresponding Author: Meghan Bradway, E-mail: Meghan.Bradway@ehealthresearch.no.

connection, communication, and the patient-provider relationship) influence the process of health behavior change. If we do not address these factors within mHealth intervention studies, we will not be able to understand the comprehensive impact of such technologies.

We propose that it is critical to design research questions that capture psycho-social factors in behavior change. Therefore, we have assessed the prevalence of questions related to these concepts in mHealth intervention studies, thereby revealing current gaps and future directions.

2. Methods

We aimed to identify articles that were published after the release of the 2015 Guidance for Industry and Food and Drug Administration Staff regarding how to address mHealth technologies [2]. These articles would thereby reflect the most updated efforts to assess new mHealth technologies, including those that address these new guidelines. We reviewed studies published in English between Jan. 1, 2015 and Jan. 18, 2019, describing mHealth interventions for patient self-management of WHO’s listed major chronic non- communicable diseases (NCD) [3], as well as chronic mental illnesses. We searched Medline, PubMed, Google Scholar, and ProQuest Research Library for combinations of “mobile application” or wearable, and self-management or self-efficacy, and patient. We focused on qualitative questions asked in the following methods: study-specific questionnaires, interviews, and focus groups. Methods that described the purpose of the inquiry, e.g. satisfaction, without listing the questions themselves, were also included. Questions asked to both patients and HCPs were then grouped under emergent themes and then overarching categories: user experiences and four major psycho-social theories of behavior change: behavior change intentions, facilitators/barriers, measures of behavior change.

3. Results

The search resulted in 31 articles. Twenty-four articles included qualitative questions (Figure 1).

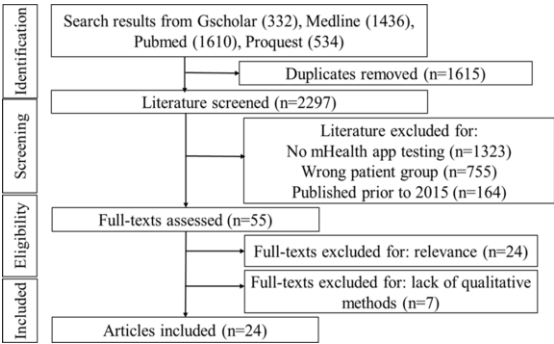


Figure 1 PRISMA diagram describing the selection of articles for data synthesis.

Emergent categories (n=18) were identified, and then grouped under broader categories (n=4).

We identified 204 questions and six articles that did not list their questions but, instead, described the topic of their inquiries. *User experience* was the most represented category (n=103 questions in n=13 articles, and n=5 articles addressing this category). Other inquiries focused on motivation, goals, and control [22], daily or SM habits [23, 24], confidence in future use [25] and focus on intention of use [26]. It is important to note that articles cited under a category may contain few questions that address that category, e.g. Fortuna et al. only included one question that addressed the category *Facilitators and barriers* [4]. Two of the seven articles that used interviews included questions that expanded upon previous feedback, e.g. “Anything else?”, “What makes you say that?”. While it was most common that patients were the target of inquiries (n=192 questions addressed to patients), three studies queried HCPs on satisfaction, experiences, and expectations (n=13 questions, n=1 interview).

Categories [Refs]	Sub-categories (n=questions)	Example questions or statements
User Experience [4-21]	<ul style="list-style-type: none"> • Experience (n=43) • Usability (n=24) • App description (n=19) • Usefulness/efficacy (n=5) • Satisfaction (n=8) • Suggested improvements (n=4) • Questions described, not listed (n=5 articles) 	<ul style="list-style-type: none"> • What were the main issues/difficulties you were facing when using the system? • The app had a remind/alert functionality • Overall, I am satisfied with how easy it is to use the App • (Nurse) Viewing daily measurements allowed nurses to recommend treatments
Measures of behavior (SM), health, or lifestyle change [5, 6, 10, 12, 22-24]	<ul style="list-style-type: none"> • SM behaviors or tasks (n=14) • General health (n=7) • General life and habits (n=6) • Mood (n=2) 	<ul style="list-style-type: none"> • The Fitbit helps me be more active • How did you feel this morning when you woke up? • How healthy do you think your diet is? • I often forget pills or take them late
Behavior change intentions [4, 5, 8, 10, 11, 15, 19, 22, 25, 26]	<ul style="list-style-type: none"> • Intentions (n=27) • Expectations (n=20) • Questions described, not listed (n=1 article) 	<ul style="list-style-type: none"> • (Provider) How confident are you that you will use the data from HeartMapp for decision making on patient care? • How likely do you think this app will help you lead a healthier lifestyle for better control of your diabetes • How comfortable would you feel sharing the information on this app with a family member or close friend?
Facilitators and barriers to behavior change [4, 6, 15, 22, 25]	<ul style="list-style-type: none"> • Self-efficacy/ autonomy (n=11) • Needs were/were not met (n=9) • App would facilitate control over SM (n=3) • Motivation (n=2) 	<ul style="list-style-type: none"> • I know what helps me stay motivated to care for my diabetes • I am able to turn my diabetes goals into a workable plan • I can ask for support for having and caring for my diabetes when I need it

4. Discussion

The number of questions related to *Behavior change intentions* and *Facilitators or barriers to behavior change* (n=72), compared to those about *User experience* and *Measures of change* (n=132), demonstrates the weight of inquiry in research toward the latter. While psycho-social factors influence the use of mHealth, this review shows that there are relatively few assessments of these forces in mHealth studies. For example, by inquiring about motivation as well as intention and external support, studies could provide a greater understanding of not just how much something has changed after a study, but also why. We need to understand the context, i.e. motivations, facilitators and mindsets, to which we are introducing mHealth interventions to understand what makes mHealth-use relevant and sustainable. Inherent factors within patients and HCPs, such as perceived roles and responsibility within chronic health care, influence how these users choose to –or not to- use an mHealth intervention. By including questions that address psycho-social factors, in addition to those that measure objective or quantitative pre-post factors, we can begin to explain when, how and why users choose to engage with mHealth in such ways that do –or do not- lead to sustainable health behavior change.

5. Conclusion

This review has demonstrated that while the qualitative questions asked in mHealth intervention studies do cover essential information, e.g. usability, there is a gap in our understanding of how and why users' choose to use mHealth interventions. By leveraging underutilized psycho-social factors, we can better understand the reasons for mHealth-use and study outcomes. Future studies could then tailor interventions to address end-user needs and more effectively optimize these technologies to facilitate health behavior change.

References

- [1] T. Song, S. Qian, T. Cui, P. Yu. The Use of Theory in Mobile Health Interventions for Patient Self-Management of Chronic Diseases. *Studies in health technology and informatics*. **264** (2019): p. 1982-3.
- [2] U.S. Food and Drug Administration. Mobile Medical Applications. Guidance for Industry and Food and Drug Administration Staff. Rockville, MD, USA: U.S. Food & Drug Administration; 2015.
- [3] Action Plan for the Prevention and Control of Noncommunicable Diseases in the WHO European Region 2016-2025. World Health Organization, <http://www.euro.who.int/en/health-topics/noncommunicable-diseases/pages/policy/publications/action-plan-for-the-prevention-and-control-of-noncommunicable-diseases-in-the-who-european-region-2016-2025>, Last access: 10.10.2019
- [4] K.L Fortuna., et al. Adapting a Psychosocial Intervention for Smartphone Delivery to Middle-Aged and Older Adults with Serious Mental Illness. *Am J Geriatr Psychiatry*. **25**(8) (2017): p. 819-28.
- [5] C. Velardo., et al. Digital health system for personalised COPD long-term management. *BMC Med Inform Decis Mak*. **17**(1) (2017) , 19.
- [6] V.G de Garibay., et al. Utility of a mHealth App for Self-Management and Education of Cardiac Diseases in Spanish Urban and Rural Areas. *J Med Syst*. **40**(8) (2016): p. 186.
- [7] Y.J Kim., et al. A Smartphone Application Significantly Improved Diabetes Self-Care Activities with High User Satisfaction. *Diabetes & metabolism journal*. **39**(3) (2015): p. 207-17.
- [8] H. Kang, H.A Park. A Mobile App for Hypertension Management Based on Clinical Practice Guidelines: Development and Deployment. *JMIR mHealth and uHealth*. **4**(1) (2016): p. e12.

- [9] N. Alnosayan, et al., Design and Usability of a Heart Failure mHealth System: A Pilot Study. *JMIR human factors*. **4**(1) (2017): p e9.
- [10] Y.J Kim, et al., A Smartphone Application Significantly Improved Diabetes Self-Care Activities with High User Satisfaction. *Diabetes & metabolism journal*. **39**(3) (2015): p. 207-17.
- [11] A. Miner, et al., Feasibility, acceptability, and potential efficacy of the PTSD Coach app: A pilot randomized controlled trial with community trauma survivors. *Psychological trauma : theory, research, practice and policy*. **8** (3) (2016): p. 384-92.
- [12] M. Munster-Segev, O. Fuerst, S.A Kaplan, A. Cahn. Incorporation of a Stress Reducing Mobile App in the Care of Patients With Type 2 Diabetes: A Prospective Study. *JMIR mHealth and uHealth*. **5**(5) (2017): p. e75.
- [13] E. Arsand, et al., Performance of the first combined smartwatch and smartphone diabetes diary application study. *Journal of diabetes science and technology*. **9**(3) (2015): p. 556-63.
- [14] J.A Naslund, K.A Aschbrenner, S.J Bartels. Wearable Devices and Smartphones for Activity Tracking Among People with Serious Mental Illness. *Mental health and physical activity*. **10** (2016): p. 10-7.
- [15] J.A Naslund, K.A Aschbrenner, L.K Barre, S.J Bartels. Feasibility of popular m-health technologies for activity tracking among individuals with serious mental illness. *Telemedicine journal and e- health : the official journal of the American Telemedicine Association*. **21**(3) (2015): p. 213-6.
- [16] J.E Owen, et al., mHealth in the Wild: Using Novel Data to Examine the Reach, Use, and Impact of PTSD Coach. *JMIR mental health*. **2**(1) (2015): p. e7.
- [17] R.M.C Masterson, et al., Review and Analysis of Existing Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self- Care Management Using the Mobile Application Rating Scale (MARS). *JMIR mHealth and uHealth*. **4**(2) (2016): p. e74.
- [18] K. Possemato, et al., Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. *General hospital psychiatry*. **38** (2016): p. 94-8.
- [19] U. Sarkar, et al., Usability of Commercially Available Mobile Applications for Diverse Patients. *Journal of general internal medicine*. **31**(12) (2016): p. 1417-26.
- [20] C.G Steele, et al., The Electronic Patient Reported Outcome Tool: Testing Usability and Feasibility of a Mobile App and Portal to Support Care for Patients With Complex Chronic Disease and Disability in Primary Care Settings. *JMIR mHealth and uHealth*. **4**(2) (2016): p. e58.
- [21] A. Triantafyllidis, et al., A personalised mobile-based home monitoring system for heart failure: The SUPPORT-HF Study. *International journal of medical informatics*. **84**(10) (2015): p. 743-53.
- [22] G Goh, et al, Short-term trajectories of use of a caloric-monitoring mobile phone app among patients with type 2 diabetes mellitus in a primary care setting. *Journal of medical Internet research*. **17**(2) (2015): p. e33.
- [23] H. Holmen, et al., Stages of change for physical activity and dietary habits in persons with type 2 diabetes included in a mobile health intervention: the Norwegian study in RENEWING HEALTH. *BMJ open diabetes research & care*. **4**(1) (2016): p. e000193.
- [24] J.Y Kim, N.E Wineinger, S.R Steinhubl. The Influence of Wireless Self-Monitoring Program on the Relationship Between Patient Activation and Health Behaviors, Medication Adherence, and Blood Pressure Levels in Hypertensive Patients: A Substudy of a Randomized Controlled Trial. *Journal of medical Internet research*. **18** (6) (2016): p. e116.
- [25] P. Athilingam, et al., Features and usability assessment of a patient-centered mobile application (HeartMapp) for self-management of heart failure. *Applied nursing research: ANR*. **32** (2016): p. 156-63.
- [26] H.S Park, H. Cho, H.S Kim. Development of Cell Phone Application for Blood Glucose Self- Monitoring Based on ISO/IEEE 11073 and HL7 CCD. *Healthcare informatics research*. **21**(2) (2015): p. 83-94.

Social Robots for Elderly Care: An Inventory of Promising Use Cases and Business Models

Somaya Ben ALLOUCH^a, and Lex van VELSEN^b

^a*Amsterdam University of Applied Sciences, Amsterdam, The Netherlands*

^b*Roessingh Research and Development, Enschede, The Netherlands*

Abstract. This paper discusses a study that aimed to elicit promising application areas and potential business models for social robotics in healthcare. For this goal, we conducted focus groups with care professionals and the management of elderly care organizations in the Netherlands and Germany. Three use cases were mentioned as the most promising: the robot as a ubiquitous aid, the robot as a helper in the room, and the robot as a guide. Finally, we discuss the implications of the medical device and privacy legislation for these three use cases.

Keywords. Social robots, healthcare, use cases, elderly care, business models, GDPR, legislation

1. Introduction

It is well known that large parts of Europe, including the Netherlands, must address growing challenges in the current health system. The number of elderly people is increasing, as is their demand for care, and Europe is struggling with a shortage of professionals in the care sector to cope with the increasing demand. To keep healthcare costs under control, it is important for older people to live independently at home for as long as possible and, where possible, to support professionals in their daily work activities to reduce the high workload. Social robots, which can be defined as robots that aim to foster “close and effective interaction with a human user for the purpose of giving assistance and achieving measurable progress in convalescence, rehabilitation, learning, etc.” [1, 2], can be of value here. For example, they can monitor the health of elderly adults and help improve or maintain elderly adults’ physical and cognitive health (e.g., by offering personalized training programmes) [3]. In addition, social robots can also be used by professionals for tasks and activities that they feel are less urgent. In this way, they can focus on the clients who need care at that moment.

The use of social robots for healthcare purposes is far from commonplace in Europe due to various issues, such as the current reimbursement system for the use of robotics in healthcare, a lack of familiarity with the possibilities of social robotics among healthcare professionals, and a lack of clarity among companies as to how they can and may use robots in healthcare institutions. This reluctance among companies is fuelled by uncertainty about privacy legislation and medical device regulations regarding robotics. To gain better insight into the role that social robotics can play in

elderly care, we created an inventory of promising use cases, potential business models and relevant regulations. To this end, we conducted a range of focus groups among care professionals and care management in the Netherlands and Germany.

2. Methods

To identify promising use cases for social robots in elderly care, we conducted three focus groups with care professionals. Two focus groups took place in the Netherlands (with six and seven participants) at care organizations that focus on elderly care in the broadest sense of the word (home care, nursing, rehabilitation, etc.). A German focus group included five participants and was held at a rehabilitation centre that treats many elderly people. The focus groups consisted of the following parts:

- Introduction of participants and the goal of the focus group;
- Moments of frustration during a workday. Here, we inventoried problematic situations about working in elderly care;
- Improvement for work activities. Here, we inventoried which solutions the participants saw for improving the problematic situations;
- Introduction of social robots. At this point, we introduced social robots and explained to the participants what they are and can generally do;
- Potential roles for social robots. Participants were asked to indicate for what type of activity (consultation, diagnosis, monitoring, training, social, lifting and carrying people or things) they thought social robots were most useful;
- Promising use cases. Together, participants looked at the problematic situations and linked these with the potential roles for social robots to identify the most promising use cases.

To uncover potential business models for the use cases we identified in the first round of focus groups and to discuss relevant legislation. We conducted focus groups with the management of elderly care organizations. Two focus groups were held with a Dutch ($n=6$ in both focus groups) institution and one with a German ($n=4$) healthcare institution. The employees who participated in this focus group consisted of Information and Communication (ICT) managers, policy-makers, training managers, innovation and care policy-makers, purchasing managers and members of the management team. During these focus groups, we presented the results of the sessions with care professionals, repeated the introduction of social robots and the inventory of potential roles, and added discussions on potential business models, implementation strategies, and applicable legislation.

3. Results

3.1. Care Professionals

During the focus groups, the participant ranked the six potential roles for social robots. They are as follows (from high to low): 1) social interaction; 2) monitoring; 3) training; 4) lifting and moving; 5) consultation; and 6) diagnosis. The professionals indicated that they would like to use the robots so that the client has “someone” to talk to because they felt that they did not always have enough time for this activity due to tight

work schedules. They also indicated that they wanted to use the robots so that they could monitor and reassure clients. A third point was that the robots could help to train the clients. A robot could perform the right exercises more often and help motivate people to perform permanent exercises so that their health could improve. It was also mentioned that robots could be used for lifting and moving clients. Finally, it was mentioned that robots could make diagnoses, but the general tendency was that people can still do this better than robots and that healthcare professionals found this function the least attractive.

An exploration of possible application areas led to three scenarios that were widely supported by the various healthcare professionals and that can already be realized with the current technology:

The robot as a ubiquitous aid. This robot records questions from clients/patients (e.g., questions about daily activities, requests for help with nausea), prioritizes the importance of these questions and only forwards a question to a professional in case of high urgency (as healthcare professionals indicate that they are overwhelmed by all the care questions they receive). The robot can also pick up objects from the ground (such as cutlery). In this way, the robot can reduce the workload of healthcare professionals and remove uncertainty from clients/patients. This use case was particularly helpful in helping clients in the common room of a care facility.

The robot as a helper in the room. This robot can look into the home of a client when an alarm goes off to determine whether an alarm is false or not. The privacy of a client is unnecessarily harmed if a professional must enter the home/room of a client at night for no reason; such a situation can be avoided by using a social robot. The robot can also set up a speech connection between the client and the healthcare professional to discuss whether help is needed. Finally, the robot can provide spoken memories (for taking medication, help with a regular daily routine). This reduces the workload of the professional, reduces unnecessary help and improves the daily rhythm of the client.

The robot as a guide. This robot reminds patients who are staying in a healthcare facility about an upcoming appointment (e.g., with a physiotherapist) and guides them to the location of this appointment (currently, patients often get lost, which is detrimental to the care provider's timetable and efficiency). In addition, the robot can guide visitors to their destination within a healthcare institution.

3.2. Management

The most important result of these focus groups was that with regard to a sustainable business case of social robots in healthcare, the majority of the participants stated that social robots should, above all, have added value for clients rather than for care professionals. The great potential that the participants saw to use social robots for patients and employees was partly reduced by many unresolved questions about the use of social robots. These questions relate to data security and liability, reliability during use by patients, quality management and integration into existing structures in the healthcare system, as well as their financing.

Interestingly, themes such as the replacement of personnel by social robots and the possible lack of human contact did not play a major role. The participants saw the possibility of relieving staff both physically and organizationally (e.g., taking over logistic tasks by a robot, completing customer surveys) and giving them more time to stay in touch with the patients.

An important barrier mentioned by participants was that they encounter problems with the implementation and use of social robotics in relation to relevant legislation and regulations, such as those that have been established for medical devices. The three user scenarios that emerged from the focus group with professionals were then used to map out relevant legislation and regulations with regard to social robots and current legislation and privacy.

3.3. Medical Device and Privacy Legislation

In the discussion with management, two main sets of legislation were discussed: the Medical Device Regulation (MDR) and the General Data Protection Regulation (GDPR). The MDR states that a device should be considered a medical device when it is used for one or more of the following purposes (limited to those that are applicable to the case of social robotics): the diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease; and/or the diagnosis of, monitoring of, treatment of, alleviation of, or compensation for an injury or disability. Key in this definition is the so-called ‘intended use’, which is stated by the manufacturer. For medical devices, the use of the device for purposes other than the intended use is prohibited, since the CE certificate is not valid when the device is not used for the intended purposes. Only in specific cases, e.g., during clinical investigations, may a device be used for purposes other than those intended, but only when this is allowed by the manufacturer.

Research into these regulations shows that robots in use cases 2 and 3 are not regarded as medical devices. After all, the robot is a means of communication (use case 2) or does not have a medical purpose (use case 3). The robot in use case 1 is regarded as a medical device, class IIa, but depending on the monitoring purposes for which the robot is used, it can also be classified as class IIb or III.

The General Data Protection Regulation (GDPR) has been in force in Europe since 25 May 2018 [4, 5]. A user of the robot (or its representative) will have to give permission for the storage and use of personal data. In addition, the supplier must provide information about the reasons for storage, how the user can influence this, and how data are exchanged. Since users will also be profiled in use case 1, the supplier will have to explain why this is done and in what way and will have to give the user the option not to use these profiling services. For use cases 2 and 3, the same rules apply (although the rules on profiling will probably not apply).

4. Discussion

In this study, we have identified the most promising use case for using social robots in elderly care and have explored potential business models. These use cases are the robot as an ever-present helper, the robot as an aid in the room, and the robot as a guide. Next, we mapped potential business cases with elderly care management and discussed related legislation. This study is one of the first to take such an approach towards social robots in elderly care or even healthcare. However, insights such as these are instrumental for developing social robotics applications, implementation plans, and exploitation plans and, as a result, are a prerequisite for successful and durable implementation.

As in any study, ours has some limitations. First, the identified use cases and insights regarding potential business models are highly influenced by the Dutch and German context and care systems. Therefore, caution should be exercised when generalizing these results beyond these two countries. Finally, elderly care has an extremely wide range of functions, involved professionals, and care protocols. To cover all of them in a study such as this one is impossible. It is possible that we missed a potential use case. However, we feel that we have generally identified the most important use cases.

The use of social robotics can be of tremendous added value for healthcare practice. However, we will have to experiment, deploy and collaborate more frequently and in an increasingly informed way, especially with professionals and clients, to determine the contexts in which the use of social robots in the healthcare sector will be most sustainable.

References

- [1] D. Feil-Seifer, M.J. Mataric, Defining socially assistive robotics. Paper presented at the 9th International Conference on Rehabilitation Robotics, 28 June-1 July 2005. <https://doi.org/10.1109/icorr.2005.1501143>.
- [2] M.M.A. De Graaf, S. Ben Allouch, and J.A.G.M. Van Dijk, Long-term evaluation of a social Robot in Real Homes, in: AISB 2014 - 50th Annu. Conv. AISB, 2014. doi:10.1075/is.17.3.08deg.
- [3] R.A.C.M. Olde Keizer, L. van Velsen, M. Moncharmont, B. Riche, N. Ammour, S. Del Signore, G. Zia, H. Hermens, and A. N'Dja, Using socially assistive robots for monitoring and preventing frailty among older adults: a study on usability and user experience challenges, *Health Technol. (Berl)*. (2019). doi:10.1007/s12553-019-00320-9
- [4] REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC
- [5] Guidance document – Classification of Medical Devices (MEDDEV 2.4/1 rev.9)

The Digital Transformation of Patient-Reported Outcomes' (PROs) Functionality Within Healthcare

Jeppe ERIKSEN^{a,1}, Pernille BERTELSEN^b, Ann BYGHOLM^c

^{a,b}Department of Development and Planning, Aalborg University

^cDepartment of Communication and Psychology, Aalborg University

Abstract. This paper elucidates how the functionality of Patient-Reported Outcome (PRO) has evolved due to its digital transformation. Hence, PROs traditional use within healthcare is described and compared to the application of electronic PROs (ePROs); leading to a discussion regarding PROs functionality. The literature included in this paper stems from a systematic scoping review. The digitalization supplements former functionalities of PRO by enabling timely, accessible, systematical and progression oriented data; however, further improvements are necessary to enhance PROs application in current healthcare.

Keywords. Patient-reported outcome, PRO, digital mediation, ePRO and PROs functionality

1. Introduction

Patient-reported outcomes (PROs) are increasingly used in current healthcare due to their multidimensional functionality, among other things, made possible by the digitalization of PRO. An alteration in the use and mediation of PROs, which have shown to be feasible [2], also as part of clinical practice [1].

A commonly used definition of PRO was coined by the American Food and Drug Administration (FDA) in 2009, stating that PRO is: “Any report of the status of patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else” [3].

Traditionally, PRO-questionnaires was used to test the effect of new drugs on patients, and as aggregated data targeted research purposes [5]. However, more extensive use of PRO in clinical practice is currently gaining ground within healthcare. Thus, data is no longer preserved clinicians and scientist, since PRO-data now is expected to inform, include and actively engage patients in managing their own disease [6]. Therefore, PROs integration into clinical practice has altered its current purpose and functionality [7].

Hence, this paper provides an overview of PROs functionality pre and post its digitalization and discuss ongoing challenges.

2. Method

The literature included in this article stems from a systematic scoping review regarding the link between PRO and patient participation. A comprehensive review,

¹ Corresponding Author: je@plan.aau.dk.

which therefore also provides an overview of PROs functionalities and digital transformation. Hence, the articles in this paper are a selective extraction, included if they pertain to PROs functionality pre or post its digital mediation.

The publications were identified using the following databases: PubMed, Embase, CINAHL and Scopus. The objective was to identify associations between patient participation and PRO. When conducting the search, the word 'patient-reported outcome' was included every time, and then combined with search words covering either: 'patient recognition', 'patient participation' or 'patient empowerment'. Exclusion criteria pertained to studies regarding primary care, traditional research studies where PRO was used as a secondary endpoint and studies where children were the subject field. Inclusion criteria consisted of articles covering, e.g. PROs used at hospitals or outpatient clinics, articles examining PROs effects, directly as well as indirectly. This strategy resulted in 6,895 articles; decreasing to 4,343 after duplicates had been removed. After reading and sorting the abstracts 256 articles were fully read and assessed for eligibility [8]; hereof, 27 articles are included in this article.

3. Applications of PRO

3.1. PROs functionality

Originally, PRO was used for group-level research and medical research, as supplement data/secondary outcome measures in clinical trials, e.g. in randomized controlled trials (RCT) studies, testing new drugs and evaluating the comparative effectiveness of healthcare interventions. Thus, the collection of PRO-data established a systematic approach when eliciting patients perspective on what mattered to them regarding their health status, their experiences pertaining to adverse effects, side-effects, comorbidity, and the burden of illness or treatment [5][9].

Therefore, the integration of PRO means that the evaluation of treatment no longer merely is based on objective criteria and physicians assessments, but on patients' perceptions and experiences as well [4]. A shift of focus in healthcare interventions, emphasizing the importance of patients health-related quality of life (HRQoL) to complement more traditional survival measures [10].

3.2. ePROs functionality

Furthermore, the digital transformation of PRO may enable new functionalities.

Firstly, when PRO is digitally mediated it allows real-time monitoring of patients and timely-data becomes available during patient-clinician consultations; potentially, leading to tailored patient management, improved decision making, more accurate diagnosing and better treatment [11].

Secondly, accessibility of PRO-data changes when digitalized as it enables patients to access data from at home; allowing them to follow their disease progression systematically over time. Therefore, the completion of ePRO-questionnaires potentially initiates a learning process as patients are confronted with issues pertaining to their condition; thus, PRO-data might empower and engage patients [12]. This feature, coupled with patients opportunity to track their disease progression over time and, e.g. set disease-related goals, are focal reasons why PRO-data is perceived as a self-management tool [13]. Thus, ePROs increased accessibility potentially improves patient

participation and self-monitoring as part of patients everyday life [13][14]. Moreover, ePROs might also affect patients’ ability to engage more actively in conversations with healthcare professionals; potentially, enhancing shared-decision making in clinical practice [11].

Furthermore, improved accessibility also means that PRO can be used as a coordination tool as it supports organizational information flows on both a vertical and horizontal level. In other words, the digitalization of PRO potentially enables healthcare teams and providers on different organizational levels to share data. As a result, stakeholders are able to base decisions on timely data, while patients’ avoid having to complete similar questionnaires over and over again, as they progress through the healthcare system [7]. However, this feature of PRO-data arguably relies heavily on adequate digital infrastructures [15] and regulations like, e.g. GDPR.

Thirdly, the systematic digital collection of PRO-data offers stakeholders an opportunity to plan and execute healthcare based on population data; exemplified by the value-based healthcare (VBHC) approach, where the reduction of health care costs and efficient use of resources are central aspects [16]. Thus, VBHC, via the use of digital PRO-data, might enable benchmarking, reimbursement and accountability of healthcare providers based on values pertaining to patient preferences and economic efficiency [14]. Moreover, this application of PRO-data might be used to improve systematic knowledge sharing among healthcare providers [14].

Lastly, PRO, as part of a triage system, is an example of a more recent application. The system is based on algorithms, which sorts patients into different categories (green, yellow and red) according to their health status; subsequently, determining who needs to show up for a patient-clinician consultation. Hence, some patients are able to skip unnecessary routine consultations, which ensures that those patients who are in need of medical attention, are prioritized [17].

Table 1. PROs functionality on an individual and population level pre and post its digitalization.

Stakeholders ²	Functionality	Pre digitalization (PRO)	Post digitalization (ePRO)
C	Decision-making and treatment [11]	x	x
C	Diagnosing [18]	x	x
CP	Patient perspective [4][6]	x	x
CP	Shared-decision making (SDM) [6]	x	x
CI	Drug testing [19]	x	x
CI	Research [14]	x	x
CPM	Patient-centred healthcare [14][20]	x	x
CP	Communication/dialogue [13][20]	x	x
C	Screening [11][20]	x	x
CPM	Patient satisfaction [6]	x	x
P	Patient participation [6]		x
P	Self-management [6][13][14][17]		x
M	Health policy development [21]		x
CPM	Quality of care [1][14][20]		x
CM	Best practice [14]		x
CP	Adherence [6][15]		x
M	Reduce health care costs/efficient use of resources [22]		x
CM	Triage system based on algorithms [17]		x

² Stakeholder terminology: C – clinicians, P – patients, M – managers/politicians, I -industry

P	Patients goals [15]	x
M	Monitor population health/preventive tool [7]	x
C	Monitoring and patient management [22]	x
M	Value-based healthcare (VBHC)/ Benchmarking/Reimbursement/ Accountability [16]	x
CM	Coordination tool/ Interdisciplinary/multidisciplinary communication [20]	x
P	Patient Empowerment [23]	x
P	Self-monitoring [24]	x

Table 1. Demonstrates how the digitalization of PROs has altered its functionality. Thus, PRO has changed from mainly being a tool used by clinicians and industry for research and drug testing, to now also being a tool used by patients, healthcare managers and politicians, e.g. facilitating systematic patient participation, patient management and VBHC.

4. Discussion

The paper elucidates how the digitalization of PROs has changed their functionality and purpose. However, even though the digitalization has improved and innovated PROs functionality, certain issues pertaining to the application and implementation of PROs still persist.

Therefore, even when ePROs, in some cases, might increase data accessibility, the use of PROs in healthcare can also exclude particular groups of patients. Thus, low health/eHealth literacy [25][26] and lack of technological access [19][22][27], are factors which potentially exclude patients; subsequently, creating a barrier towards patient participation. Hence, clinicians and managers need to be aware of the patient burden PROs might entail [22][27], to make sure that the use of PROs enables patient empowerment and not turns into patient work [28].

Furthermore, it is essential to make sure that PRO-data is actionable [6] and actually used in clinical practice since this is not always the case [29]. Therefore, to secure the use of PRO-data in clinical practice PRO measures need to be clinically relevant [22][29], and clinicians should be trained regarding the application and interpretation of PRO-data [29][30]. Whereas, on an organizational level, attention to how workflows are affected [27] and whether sufficient time is allocated for meaningful use of PROs [30] is needed.

Hence, even after PROs digital transformation, there is still a lot of work to be done to make sure that PRO fulfil its potential, which is why further research within this area is recommended. However, as the overview in table 1. Demonstrates, ePROs have the potential to further a more personalized and actively engaging approach in healthcare.

References

[1] M. Korzeniowski et al., “Piloting prostate cancer patient-reported outcomes in clinical practice,” *Support. Care Cancer*, vol. 24, no. 5, pp. 1983–1990, May 2016.

[2] D. Desantis et al., “A clinical perspective on collecting patient-reported outcomes at the point-of-care for urinary incontinence,” *Can. Urol. Assoc. J.*, vol. 10, no. 5–6, p. S87, 2016.

[3] U. S. Department of Health and Human Services, Food and Drug Administration (FDA), “Guidance for Industry Use in Medical Product Development to Support Labeling Claims Guidance for Industry,” no. December, pp. 1–39, 2009.

- [4] S. Dean, et al., “‘The patient is speaking’: Discovering the patient voice in ophthalmology,” *Br. J. Ophthalmol.*, vol. 101, no. 6, pp. 700–708, 2017.
- [5] C. A. McHorney, “HEALTH STATUS ASSESSMENT METHODS FOR ADULTS: Past Accomplishments and Future Challenges,” *Annu. Rev. Public Health*, vol. 20, no. 1, pp. 309–335, 2002.
- [6] M.-J. Santana and D. Feeny, “Framework to assess the effects of using patient-reported outcome measures in chronic care management,” *Qual. Life Res.*, vol. 23, no. 5, pp. 1505–1513, Jun. 2014.
- [7] Munch-Petersen et al., “Program PRO” VibIS, p. 70, 2016, located at: <https://danskepatienter.dk/vibis/om-vibis/vores-projekter/program-pro>.
- [8] J. Eriksen et al., “Associations between Patient-Reported Outcomes and Patient Participation - a systematic search and review.” Forthcoming, 2020.
- [9] S. P. McKenna and J. Wilburn, “Patient value: its nature, measurement, and role in real world evidence studies and outcomes-based reimbursement,” *J. Med. Econ.*, vol. 21, no. 5, pp. 474–480, 2018.
- [10] H. Fagerlinde et al., “Patients’ understanding of the concepts of health and quality of life,” *Patient Educ. Couns.*, vol. 78, no. 1, pp. 104–110, 2010.
- [11] M. Y. Halyard, “The use of real-time patient-reported outcomes and quality-of-life data in oncology clinical practice,” *Expert Rev. Pharmacoeconomics Outcomes Res.*, vol. 11, no. 5, pp. 561–570, 2011.
- [12] W.G. Groen et al., “Empowerment of Cancer Survivors Through Information Technology: An Integrative Review,” *J. Med. Internet Res.*, vol. 17, no. 11, p. e270, Nov. 2015.
- [13] C. Trillingsgaard et al., “Use of patient-reported outcomes in outpatient settings as a means of patient involvement and self-management support – a qualitative study of the patient perspective,” *Eur. J. Pers. Centered Healthc.*, vol. 4, no. 2, p. 359, 2016.
- [14] M. C. Wang and J. Bellows in M. R. Helton and T. Daaleman, *Chronic Illness Care*. 2018.
- [15] L. M. Philpot et al., “Barriers and Benefits to the Use of Patient-Reported Outcome Measures in Routine Clinical Care: A Qualitative Study,” *Am. J. Med. Qual.*, p. 1062860617745986, Dec. 2017.
- [16] M. E. Porter, “What Is Value in Health Care?,” *Perspective*, vol. 363, no. 1, pp. 1–3, 2010.
- [17] C. T. Mejdahl et al., “PRO-based follow-up as a means of self-management support – an interpretive description of the patient perspective,” *J. Patient-Reported Outcomes*, vol. 2, no. 1, p. 38, 2018.
- [18] K. Haywood et al., “Patient participation in the consultation process: A structured review of intervention strategies,” *Patient Educ. Couns.*, vol. 63, no. 1–2, pp. 12–23, 2006.
- [19] L.E. Strong, “The past, present, and future of patient-reported outcomes in oncology,” *Am. Soc. Clin. Oncol. Educ. Book*, pp. e616–e620, 2015.
- [20] J. Greenhalgh, “The applications of PROs in clinical practice: What are they, do they work, and why?,” *Qual. Life Res.*, vol. 18, no. 1, pp. 115–123, 2009.
- [21] N. Black, “Patient reported outcome measures could help transform healthcare,” *BMJ*, vol. 346, no. 7896, pp. 1–5, 2013.
- [22] C. Chang, “Patient-reported outcomes measurement and management with innovative methodologies and technologies,” *Qual. Life Res.*, vol. 16, no. SUPPL. 1, pp. 157–166, 2007.
- [23] M. McAllistar et al., “Patient empowerment: the need to consider it as a measurable patient-reported outcome for chronic conditions,” *BMC Health Serv. Res.*, vol. 12, no. 1, p. 157, Jun. 2012.
- [24] L. Segan et al., “Enhancing self-care strategies in heart failure through patient-reported outcome measures,” *Intern. Med. J.*, vol. 48, no. 8, pp. 995–998, Aug. 2018.
- [25] J. Øvretveit et al., “Using patient-reported outcome measurement to improve patient care,” *Int. J. Qual. Heal. Care*, vol. 29, no. 6, pp. 874–879, 2017.
- [26] J. S. Blumenthal-Barby, “‘That’s the doctor’s job’: Overcoming patient reluctance to be involved in medical decision making,” *Patient Educ. Couns.*, vol. 100, no. 1, pp. 14–17, 2017.
- [27] S. E. Tevis et al., “Patient-Reported Outcomes for Breast Cancer,” *Ann. Surg. Oncol.*, vol. 25, no. 10, pp. 2839–2845, 2018.
- [28] A. Bygholm and P. Bertelsen, “Patient at Work - In the Era of PRO.,” *Stud. Health Technol. Inform.*, vol. 241, pp. 115–121, 2017.
- [29] J. Greenhalgh et al., “How do doctors refer to patient-reported outcome measures (PROMS) in oncology consultations?,” *Qual. Life Res.*, vol. 22, no. 5, pp. 939–950, Jun. 2013.
- [30] D. Howell et al., “Patient-reported outcomes in routine cancer clinical practice: A scoping review of use, impact on health outcomes, and implementation factors,” *Ann. Oncol.*, vol. 26, no. 9, pp. 1846–1858, Sep. 2015.

The Effect of Patient Accessible Electronic Health Records on Communication and Involvement in Care - A National Patient Survey in Sweden

Jonas MOLL^{a,1} and Hanife REXHEPI^b

^a*Centre for empirical research on information systems, Örebro University, Sweden*

^b*Department of Information Technology, University of Skövde, Sweden*

Abstract. During recent years, patient accessible electronic health records (PAEHRs) have been implemented nationally in Sweden, as well as internationally, as a means of supporting patient engagement and shared decision making. Few studies have, however, investigated the long-term effects of PAEHRs on communication with care professionals and involvement in care. The national survey study presented here, answered by 2587 patients in Sweden, focuses on these aspects specifically. The results show that the Swedish PAEHR system *Journalen* has had a positive impact on communication with care overall (84% agree or strongly agree with that communication with medical staff has improved), but only 31% agree or strongly agree with that the content of the PAEHR is discussed with care professionals. *Journalen* also seems to have a positive impact on involvement in care, but the results are mixed when it comes to effects on shared decision making.

Keywords. eHealth, patient accessible electronic health records, communication, patient involvement, national survey

1. Introduction

The communication between clinicians and their patients is a fundamental component of clinical practice and is essential for achieving quality healthcare [1][2]. In recent years, a growing emphasis on patient autonomy, patient-centered care and patient participation has further exemplified the importance of effective doctor-patient communication [1]. Prior research on patient complaints has shown that poor attitudes and poor quality of information provided are the most commonly identified communication errors made by doctors [1]. In Sweden, a new legislation - the Patient Act or Patientlagen - was approved in 2015 with the aim to strengthen the patient's position and rights in healthcare. The patient act covers areas such as accessibility, patient participation, information provided by healthcare professionals and consent. However, studies have reported that the act has had limited impact on healthcare [3]. Patients are still feeling poorly informed regarding their

¹Corresponding Author: Jonas Moll, E-mail: jonas.moll@oru.se.

test results, treatment options and management and less involved in their own healthcare [3]. These results are worrying as there is a growing recognition that patients can participate more actively in decisions regarding their own care when they are provided with information [4] [5]. Moreover, the WHO in their policy framework for Health 2020 [6] and the Swedish government [7], like other governments worldwide e.g. England [8], advocate and endorse the need for healthcare organizations to support patients being active participants in their own care. Therefore, a lot of effort has been put into the development of new technologies for health (eHealth) with tremendous potential for promoting patient participation, enhancing information-sharing and communication between patients and healthcare professionals and improving health outcomes [9]. An example of this type of a service is to provide patients with easy access to their electronic healthcare record (EHR), so called patient accessible EHR (PAEHR), through secure eHealth services. The overall goal of PAEHR is to improve communication between healthcare professionals and patients, ultimately increasing patients' participation in care and improving the shared decision-making process [10]. Two of the most well-known examples of PAEHR are OpenNotes in USA [5] [11] and MyUHN Patient Portal in Canada [12]. A similar service was launched in 2012 in Sweden, when Region Uppsala gave all citizens over 18 years of age online access to their EHR through the eHealth service *Journalen*. This service was eventually launched as a national system in Sweden, accessible through a national patient portal. However, PAEHR are still accompanied by major concerns from healthcare professionals, who stress that online access might lead to an increased workload, privacy risks and cause patients to worry [13] [14]. Studying and understanding if and how PAEHR affects communication between healthcare professionals and patients is important as communication is essential for the quality of healthcare. Moreover, patients use of and attitudes toward PAEHRs has up until today mostly focused on general attitudes to the prospective introduction of such eHealth services. Hence the aim of this paper is to investigate the effects of *Journalen* on communication with healthcare professionals and the patients' involvement in their own care.

2. Method

A national patient survey was made available through a link on the login page of the Swedish PAEHR system *Journalen*. The survey was ethically approved by the Regional Ethical Review Board in Uppsala (EPN 2016/129). The survey, which was open for all patients accessing *Journalen* during the five months period June-October 2016, included questions on several topics such as attitudes and reactions, access to and use of information, personal health, usage behavior, information literacy and information security. The questionnaire included questions with various response options (5-point Likert scale, multiple choice, and free text form). In total, 2587 patients responded to the survey, giving a response rate of 0.61% (2587/423141). Moll et al. [15] presents the study design in more depth and some overview results from the survey. Even though the results from two of the questions in this study (a and d) are also presented in the overview article (Moll et al. [15]), the contexts in which those results are presented are different - while this paper focuses on communication and involvement the overview article focuses on general usage behaviours and usage patterns. Only completed questionnaires have been analyzed, as the answers were stored in the database only when the respondent chose to

submit on the last page. This paper covers two important themes from the survey, and the debate about eHealth systems in general – namely (1) communication with healthcare professionals and (2) involvement in care. SPSS 25 was used for all analyses.

3. Results

When it comes to gender distribution among respondents, the majority (1629/2587 = 63%) of the respondents were women, which corresponds well with official data on usage of the system. The majority of the respondents (61 %) also had an educational background of at least three years in higher education.

3.1. Communication with clinicians

Results related to questions on communication are presented in Table 1. The questions/statements are: (a) *The PAEHR improves communication between medical staff and me*; (b) *To take part of the patient information via Journalen has affected the relationship with health care system positively*; and (c) *I discuss the content of Journalen with medical staff*. As can be seen from the table, the majority of the respondents believe that access to *Journalen* has improved the communication with medical staff and the relationship with healthcare overall. On the other hand, the content of *Journalen* does not seem to be discussed with medical staff to any large extent.

Table 1. Results regarding PAEHR in relation to communication. The highest value for each question is indicated with bold face.

Question	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
a	1418 (56%)	705 (28%)	280 (11%)	67 (3%)	55 (2%)
b	865 (34%)	781 (31%)	656 (26%)	95 (4%)	118 (5%)
c	255 (10%)	511 (21%)	448 (18%)	329 (13%)	945 (38%)

3.2. Involvement in care

Results related to questions on involvement in care are presented in table 2. The questions/statements are: (d) *I use Journalen to become more involved in my care*; (e) *Information in Journalen had a positive impact on the ability to work together with medical staff making decisions about care and treatment*; (f) *Information in Journalen had a positive impact on the ability to follow the prescription of treatment*; and (g) *Information in Journalen had a positive impact on the ability to take own steps to improve health*. As can be seen in the table the majority of the respondents strongly agree or agree with all statements, even though there is a rather high portion of neutral respondents when it comes to shared decision making and self-care.

4. Discussion and Conclusions

The results presented in this paper show how patients experience that their communication with healthcare professionals and their involvement in the care process has been

Table 2. Results regarding PAEHR in relation to involvement in care. The highest value for each question is indicated with bold face.

Question	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
d	1372 (56%)	651 (26%)	283 (12%)	57 (2%)	107 (4%)
e	571 (24%)	645 (27%)	785 (32%)	186 (8%)	215 (9%)
f	848 (35%)	712 (29%)	575 (24%)	117 (5%)	170 (7%)
g	652 (27%)	681 (28%)	725 (30%)	139 (6%)	207 (9%)

affected by PAEHR. The results are important for our understanding of the effects of PAEHR for patients, since they are based on the first large follow-up study conducted several years after launch of *Journalen* in Sweden. Earlier studies, like the interview study with cancer patients conducted by Rexhepi et al. [16] one year after launch, rather reflects short-term effects. It is clear that the majority of the respondents believe that access to information from PAEHR allows them to communicate more effectively in conversations with healthcare professionals regarding their own care. One explanation for these results may be that access to test results and information about care plan, treatments etc from PAEHR can, as reported in studies such as [17] and [18], improve the patients' understanding of their medical condition. Hence, it is possible that the patients' understanding of their illness affect their ability to engage in meaningful conversations with the physician. Moreover, as a result of accessing information from the PAEHR, the majority of the patients also report an improved communication and relationship with healthcare professionals.

Furthermore, a significant amount of patients reported that they use *Journalen* to become involved in their own care. However, when it comes to involvement in terms of increased ability to engage in self-care and shared decision making, the results are not as positive. The reasons are difficult to find in a survey, however engaging in e.g., shared decision making is a two-way-process where a clinician and patient jointly participate in making a medical decision [19]. If patients do feel that their opinions are not valued, they may be less inclined to engage in the meaningful conversations needed for effective shared decision making. Hence, provision of information alone from PAEHR will not automatically result in patients engaging in shared decision making. Instead, healthcare professionals must incorporate patients into the decision-making process. Future studies will be needed to look more closely at the association between note reading and shared decision making. Furthermore, despite ethical principles and legal requirements to engage patients in their own care, some patients may not want to be active participants. It is however vital that clinicians assess and understand the role each patient wishes to play in his or her care. Moreover, the majority of the respondents also strongly agree with that *Journalen* makes it easier for them to follow prescriptions, which is in line with earlier studies in the area [20]. The results also show that the contents of the PAEHR is not discussed to any large extent during the patient meeting. This is not surprising considering the resistance shown by the healthcare professionals, especially physicians and their trade union, in Sweden [14]. However, given that patients, over all, perceive PAEHR as positive and important it is of high relevance to study why the contents of PAEHR are not discussed with patients and how they can be integrated into the patient meeting.

References

- [1] Kee, J. W., Khoo, H. S., Lim, I., & Koh, M. Y. Communication skills in patient-doctor interactions: learning from patient complaints, *Health Professions Education* **4**(2) (2018), 97-106.
- [2] Spooner, K. K., Salemi, J. L., Salihu, H. M., & Zoorob, R. J. Disparities in perceived patient-provider communication quality in the United States: Trends and correlates, *Patient education and counseling* **99**(5) (2016), 844-854.
- [3] Vårdanalys. *Lag utan genomslag* (2017:2). Available at: <https://www.vardanalys.se/rapporter/lag-utan-genomslag/> [Accessed: 2019-01-29], in Swedish.
- [4] Osei-Frimpong, K., Wilson, A., & Lemke, F. Patient co-creation activities in healthcare service delivery at the micro level: The influence of online access to healthcare information, *Technological Forecasting and Social Change* **126** (2018), 14-27.
- [5] Bourgeois, F., Lowe, E., & Wachenheim, D. (2018). OpenNotes: Sharing Visit Notes With Patients and Families, *Pediatric Nursing* **44**(1) (2018), 45-48.
- [6] WHO. *Health 2020 A European policy framework and strategy for the 21st century* (2013). Available at: <http://www.euro.who.int/en/publications/abstracts/health-2020.-a-european-policy-framework-and-strategy-for-the-21st-century-2013> [Accessed 2019-01-29].
- [7] Regeringen. *Vision E-hälsa 2025* (2016). Available at: <https://www.regeringen.se/information/material/2016/04/vision-e-halsa-2025/> [Accessed 2019-01-20], in Swedish.
- [8] National Clinical Guideline Centre. *Patient Experience in Adult NHS Services: Improving the Experience of Care for People Using Adult NHS Services: Patient Experience in Generic Terms* (2016).
- [9] Barelo, S., Triberti, S., Graffigna, G., Libreri, C., Serino, S., Hibbard, J., & Riva, G. eHealth for Patient Engagement: A Systematic Review, *Frontiers in psychology* **6** (2016), 2013.
- [10] Safford, M. M. A. New Chapter in Patient-Centered Care: Sharing the Medical Note?, *Ann Intern Med* **168**(4) (2018), 298-298.
- [11] Walker, J., Meltsner, M., & Delbanco, T. US experience with doctors and patients sharing clinical notes, *Bmj* **350** (2015).
- [12] Wiljer, D., Urowitz, S., Apatu, E., et al. Patient accessible electronic health records: exploring recommendations for successful implementation strategies, *J Med Internet Res* **10**(4) (2008), e34.
- [13] de Lusignan, S., Mold, F., Sheikh, A., Majeed, A., Wyatt, J. C., Quinn, T., ... & Blakey, H. Patients' online access to their electronic health records and linked online services: a systematic interpretative review, *BMJ open* **4**(9) (2014), e006021.
- [14] Grünloh, C., Myretg, G., Cajander, ., & Rexhepi, H. Why do they need to check me? patient participation through eHealth and the doctor-patient relationship: qualitative study, *Journal of medical Internet research* **20**(1) (2018), e11.
- [15] Moll, J., Rexhepi, H., Cajander, Å., Grünloh, C., Huvila, I., Häggglund, M., ... & Åhlfeldt, R. M. Patients' experiences of accessing their electronic health records: national patient survey in Sweden, *Journal of medical Internet research* **20**(11) (2018), e278.
- [16] Rexhepi, H., Åhlfeldt, R. M., Cajander, Å., & Huvila, I. Cancer patients' attitudes and experiences of online access to their electronic medical records: A qualitative study, *Health informatics journal* **24**(2) (2018), 115-124.
- [17] Delbanco, T., Walker, J., Bell, S. K., Darer, J. D., Elmore, J. G., Farag, N., ... & Ross, S. E. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead, *Annals of internal medicine* **157**(7) (2012), 461-470.
- [18] Bell, S. K., Gerard, M., Fossa, A., Delbanco, T., Folcarelli, P. H., Sands, K. E., ... & Walker, J. A patient feedback reporting tool for OpenNotes: implications for patient-clinician safety and quality partnerships, *BMJ Qual Saf* **26**(4) (2017), 312-322.
- [19] Elwyn, G., Frosch, D., Thomson, R., Joseph-Williams, N., Lloyd, A., Kinnnersley, P., ... & Edwards, A. Shared decision making: a model for clinical practice, *Journal of general internal medicine* **27**(10) (2012), 1361-1367.
- [20] Wright, E., Darer, J., Tang, X., Thompson, J., Tusing, L., Fossa, A., ... & Walker, J. Sharing physician notes through an electronic portal is associated with improved medication adherence: quasi-experimental study, *Journal of medical Internet research* **17**(10) (2015).

The Patient as Genomic Data Manager – Evaluation of the PROMISE App

Lena GRIEBEL^{a,1}, Marc HINDERER^a, Ali AMR^b, Benjamin MEDER^b, Marc Schweig^c, Dominic Deuber^d, Christoph Egger^d, Claudia Kawohl^e, Annika Krämer^e, Isabell Flade^f, Dominique Schröder^d, Hans-Ulrich PROKOSCH^a

^a Chair of Medical Informatics, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen

^b Department of Medicine III, University of Heidelberg, Heidelberg

^c Backes: SRT GmbH, Saarbrücken

^d Chair of Applied Cryptography, Friedrich-Alexander-University Erlangen-Nürnberg, Nürnberg

^e CISA Helmholtz Center for Information Security, Saarbrücken

^f CeGaT GmbH, Tübingen

Abstract. PROMISE (Personal Medical Safe) was a German research project which aimed to provide the responsibility of genomic data to the patient via a mobile app. The patient should accept or decline study requests to use his/her genomic data via the app. In the evaluation of the app the experiences with mobile health as well as the opinion on being the genomic data manager were measured. Furthermore, the test patients were asked about their opinion and their concerns on the PROMISE app. Most of the 19 test patients were aware of the high sensibility of genomic data and thought that the PROMISE app was a suitable solution. The largest part found it good that they were the responsible data owner. However, several participants also found it important to have a permanent contact person when it comes to questions on inquiries or the app.

Keywords. Personalized medicine, genomic sequencing, patient empowerment, technology acceptance

1. Introduction

In today's healthcare there is a shift to empower the patient from being a mere recipient of medical treatments to being an active partner in the own healthcare process [1]. It has been shown that patients that are willing to actively participate potentially can achieve better health outcomes than more passive patients [2-4]. In Germany statutory health insurances are obligated to provide electronic health records for their patients starting in the year 2021; and consumer-based health apps can be prescribed by physicians to their patients in the future [6].

A big trend in current healthcare is the so-called personalized medicine which takes individual factors into account and offers therapeutic measures individually tailored for each specific patient [7]. This includes the use of genomic data to develop

1 Corresponding Author: E-mail: lena.griebel@fau.de

the best therapy options. Different institutions also access and use anonymized genomic patient data for research purposes. Until now patients however mostly have no or very limited access to their genomic data which have been collected in the clinical setting. In the German BMBF funded (German Federal Ministry of Education and Research) research project PROMISE DS (Personal Medical Safe) the patient should be an active partner in this genomic medicine assuming that he or she is indeed able and willing to. Studies have shown that patients in general want to be involved in new health technologies [8, 9] but that it is necessary to assess the patients' preferences towards engagement strategies beforehand [10].

PROMISE DS developed and provided an infrastructure and patient app, which allows encrypted storage of genomic data in the background and patient management of access to this data in the foreground. The objective of our companion study was to evaluate patients' attitude towards and acceptance of this innovative technology. We wanted to gain insights into the potential genomic data managers' point of view and thus learn for the further development and deployment of such a new and innovative technology. The main research questions were:

1. How much experiences have the patients with apps and health apps?
2. What is the knowledge of the patient regarding genomic testing and what are the patients' concerns?
3. What do patients think of the PROMISE app?
4. What is the patients' general opinion on being the genetic data manager?

2. Methods

We selected test patients to use the PROMISE app over four months. All the test patients were cardiomyopathy patients recruited at Heidelberg university hospital. Cardiomyopathy is a disease known for its genetic predisposition. The genomes of all patients in our study have been sequenced before.

We included patients with a minimum age of 18 years with a medical indication for genetic testing and regular access to an Android-based smartphone. Patients who fulfilled the inclusion criteria were given an information sheet about the study and had to give their informed consent. The treating physician helped them to install the PROMISE app on their personal smartphone. It was made clear that the app would not offer medical advantages for the patient but the focus of the project was to learn about possible ways of designing an app with the patient as genomic data manager. Over the course of four months the patients received 12 simulated queries through the mobile application. There were no real customers; the requests have been developed by project members. The inquiries were designed to look as if they originated from hospital researchers, university researchers and commercial companies. The patient had the options 1) to accept, 2) to decline or 3) to ignore the request. Before and after the use of the PROMISE app the patients each fulfilled a questionnaire consisting out of closed questions in a pen-and-paper-approach. The questionnaires were designed on the basis of self-developed questions and already existing tools such as the questionnaire on patients' perspective on pharmacogenetics testing by Rogausch et al [11]. Supplementary to the paper-based questionnaires interviews were performed at the second time point to further assess the opinion toward genetic testing. The evaluation

was performed with 19 patients at both time points – the interviews with 11 participants.

3. Results

3.1 App Using Experience and Knowledge on Genomic Testing

We asked the participants about their current use of apps and health apps.

The largest part said that they used apps on a daily basis (14 of 19). Nevertheless, ten of 19 participants stated that they never use specific health apps, five used health apps very rarely and only each two persons stated to use health apps on a daily or weekly basis.

The test patients were asked to self-assess their knowledge regarding technical terms such as “gene”, “genetics”, “genome” or “pharmacogenomics”. Whereas the knowledge on the terms “gene” and “genetics” was rather good, the patients assessed their knowledge on “pharmacogenomics” very low. 15 out of 19 test patients said after the test use of the PROMISE app that they are aware or very aware that genetic data are particularly sensitive; before they used the app also 13 patients agreed or fully agreed on that. In the interviews data security was as well revealed as a major concern regarding genetic testing.

Also the major part of the participants agreed or fully agreed that genetic information should only be stored when high safety precautions are met (15 agreed/fully agreed before the test use; 18 agreed/fully agreed after the test use). Most participants (14 before the test use, 18 after the test use) stated that they thought that the safety concept of PROMISE protected their privacy well or very well.

3.2 Opinion on the PROMISE App

After the test use of PROMISE most participants said that they found the (simulated) study inquiries very interesting or interesting (79%), 21% were unsure about it. None of them did find the inquiries disruptive in their daily life.

Also after testing PROMISE, 10 persons were afraid to falsely accept inquiries due to misunderstanding them, only three to falsely decline inquiries. In the interviews nearly every participant stated that the PROMISE app was a good idea. Several participants also mentioned that it was very important to have a permanent contact person to ask when it comes to difficulties or uncertainties when using the PROMISE app.

3.3 Opinion on Being the Data Manager

Most of the test persons found it good or very good (15 before and 18 after the test use) to own a safely stored copy of their own genetic data. Before the test use four persons were unsure, after the test use only one person was unsure about this question. Also in both time points most participants stated that they were satisfied or highly satisfied with being in control over the use of the genetic sequence data (13 before, 17 after).

Nevertheless, 5 resp. 4 persons said that they were afraid of being the responsible person for the genetic sequence data.

4. Discussion

Even though the participants stated that they used apps in general very frequently, almost no one used health apps frequently. The general self-assessed knowledge on genomics and related terms was relatively low. Most test persons saw genetic data as especially sensitive and in need of high protection but also found the security concept of PROMISE convincing. Most participants were interested in the simulated inquiries and found them non-disruptive. They were more afraid to falsely accept study inquiries than to falsely decline study inquiries.

Owning a personal copy of the genomic data was found good or even very good by most of the participants. Also, most persons were satisfied with being in control over the use of the data. Besides the fact that the largest part of participants was not afraid of it, it has to be taken into account that several persons indeed were afraid of being the only data owner.

5. Conclusion

In our small study setting it got clear that data security plays an important role for the data-managing users. Further we might conclude that the layperson users would need support when it comes to the usage of technical/medical terms and scientific concepts in study inquiries, e.g. by a permanent contact person such as their treating physician. Also the test users were not familiar with the use of health apps. In general one might say that the users should not feel overtaxed by their responsibilities.

Although the findings of our study are not representative due to the small number of participants they nevertheless show that medical laypersons are indeed interested and willing to be the data manager of personal genetic data but that electronic services supporting them should be designed very thoroughly regarding their comprehensibility and data security and that a “real” contact person would be important.

References

- [1] G. Elwyn, D. Frosch, R. Thomson, et al. Shared Decision Making: A Model for Clinical Practice. *Journal of General Internal Medicine* **27(10)** 2012,1361-7.
- [2] J.H. Hibbard, E.R. Mahoney, R. Stock, M. Tusler. Do increases in patient activation result in improved self-management behaviors? *Health services research* **42(4)** 2007, 1443-63.
- [3] H. Holman, K. Lorig. Patients as partners in managing chronic disease. *BMJ* **320(7234)** 2000, 526-527.
- [4] D.M. Mosen, J. Schmittiel, J. Hibbard, D. Sobel, C. Remmers, J. Bellows. Is patient activation associated with outcomes of care for adults with chronic conditions? *The Journal of ambulatory care management* **30(1)** 2007,21-29.
- [5] L. Ricciardi, F. Mostashari, J. Murphy, J.G. Daniel, E.P. Siminerio. A national action plan to support consumer engagement via e-health. *Health affairs (Project Hope)* **32(2)**, 376-84.
- [6] Bundesministerium für Gesundheit. Ärzte sollen Apps verschreiben können. 2019 [updated 3.12.19; cited 2019]. Available from: <https://www.bundesgesundheitsministerium.de/digitale-versorgung-gesetz.html>.
- [7] S. Schleidgen, C. Klingler, T. Bertram, W.H. Rogowski, G. Marckmann. What is personalized medicine: sharpening a vague term based on a systematic literature review. *BMC Medical Ethics*. **14(1)** 2013, 55.
- [8] S. Safi, T. Thiessen, K.J. Schmailzl. Acceptance and Resistance of New Digital Technologies in Medicine: Qualitative Study. *JMIR research protocols*. **7(12)** 2018, e11072-e.

- [9] B. David, K. Leslie. Inside the patient journey - Three key touch points for consumer engagement strategies. Findings from the Deloitte 2018 Health Care Consumer Survey. *Deloitte*. 2018.
- [10] T. Jerofke-Owen, J. Dahlman J. Patients' perspectives on engaging in their healthcare while hospitalised. *Journal of clinical nursing*. **28(1-2)** 2019, 340-50.
- [11] A. Rogausch, D. Prause, A. Schallenberg, J. Brockmoller, W. Himmel. Patients' and physicians' perspectives on pharmacogenetic testing. *Pharmacogenomics*. **7(1)** 2006, 49-59.

Towards a Highly Usable, Mobile Electronic Platform for Patient Recruitment and Consent Management

Daniel ROBINS^{a1}, Rachel BRODY^a, In Cheol JEONG^a, Irena PARVANOV^a,
Jiazhen LIU^a, and Joseph FINKELSTEIN^a

^a*Icahn School of Medicine at Mount Sinai, New York, NY, USA*

Abstract. This study seeks to assess usability and acceptance of E-Consent on mobile devices such as tablet computers for collecting universal biobank consents. Usability inspection occurred via cognitive walkthroughs and heuristics evaluations, supplemented by surveys to capture health literacy, patient engagement, and other metrics. 17 patients of varied ages, backgrounds, and occupations participated in the study. The System Usability Scale (SUS) provided a standardized reference for usability and satisfaction, and the mean result of 84.4 placed this mobile iteration in the top 10th percentile. A semi-structured qualitative interview provided copious actionable feedback, which will inform the next iteration of this project. Overall, this implementation of the E-Consent framework on mobile devices was considered easy-to-use, satisfying, and engaging, allowing users to progress through the consent materials at their own pace. The platform has once again demonstrated high usability and high levels of user acceptance, this time in a novel setting.

Keywords. electronic consent, usability inspection, patient engagement, mobile app

1. Introduction

The E-Consent platform has matured through multiple iterations to capture patient consents for research studies in a highly usable way. Its central premise is to deliver interactive multimedia content that enhances patient engagement and education relevant to a study. Essentially, research staff just upload an approved informed consent and multimedia into the system, which then reorganizes the materials and presents it to end-users in a simple and easy-to-understand fashion. The key component of this approach, however, is the interface: the interface must be intuitive for the target audience and exceptionally easy-to-use for optimal engagement. By utilizing an iterative development approach and collecting feedback from key stakeholders and subject matter experts, our team refined the user interface over numerous cycles. The overarching goal of this multifaceted project is to prospectively assess usability, patient engagement, and acceptance of the system across a multitude of settings.

Previously we demonstrated its highly usable interface and acceptance when implemented at a workstation-like setting [1]. This study represents the continued usability inspection performed on the most recent iteration of the system, while also extending the platform to mobile devices such as tablets with touchscreens.

¹ Daniel Robins, E-mail: Daniel.Robins@mountsinai.org.

2. Methods

2.1. System Design

As described in our previous studies, system design has been guided by the Cognitive Theory of Multimedia Learning; information is structured on learning principles accounting for working memory and cognitive load [2]. The user interface was informed by our multiple previous works to deploy tools aimed at improving health literacy and patient engagement [3][4], while also accommodating limited literacy or numeracy [5]. Lessons learned from these studies have led to a Table-of-Contents design for the main menu of the application. Complex information is organized into a series of smaller ‘chapters’. Each chapter is structured similarly, beginning with a video explanation (Figure 1) of a person who explains the section in plain language. Subsequent pages in the chapter feature a few sentences of text that reiterate and expand on the relevant content. One-question quizzes appear periodically, to reinforce specific concepts. The full consent form is available at all times and on all pages of the system via the ‘Consent Now’ link. In any event, the user is automatically presented with the full consent form and the option to digitally sign it (Figure 2) upon completing all chapters.

For this study, content was entirely in American English.

2.2. Study Design

The study was designed to evaluate the usability of the system on mobile devices by eliciting quantitative and qualitative feedback.

Participants were all patients at the Mount Sinai Hospital in New York, coming from clinics including primary care, specialty care, and employee health. Upon sitting down with a tablet computer, participants were given a packet of surveys, followed by instructions to complete 3 tasks, then an additional package of evaluations and surveys to record their feedback. The tablet computer included a touchscreen, a stylus, and a keyboard with trackpad, and it was preconfigured to display the Electronic Consent system on screen.

Surveys began with a sociodemographics form and the Rapid Estimate of Adult Literacy in Medicine (REALM). Participants then performed three tasks as part of the cognitive



Figure 1. Video introduction to a chapter

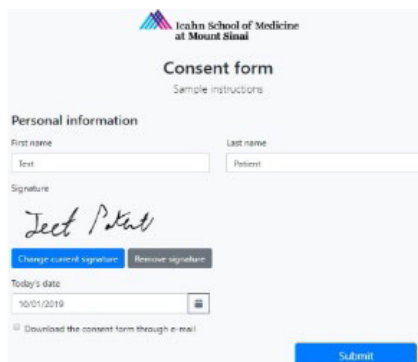


Figure 2. Signature Page

walkthrough. Afterwards, participants provided feedback on post-task questionnaires, a heuristics evaluation form, and the Process and Quality of Informed Consent for Clinical Research (P-QIC) form. A semi-structured qualitative exit interview completed the session.

The three tasks for the cognitive walkthrough could be completed with or without assistance from the research associate; if a participant needed help, this request was recorded. Task 1 instructed the user to complete all chapters of the educational module. Task 2 required users to complete the consent module and electronically sign the consent form. Task 3 asked users to complete an electronic version of the System Usability Scale (SUS).

After completing tasks 1 and 2, the participant was asked to grade each task on a scale of 1 (very difficult) to 5 (very easy) using a survey that included the following questions: 1) How difficult or easy was it to review the content and finish the sections? 2) How difficult or easy was it to answer the questions? 3) How satisfied are you with using this system to complete this task? 4) How would you rate the amount of time it took to complete this task? 5) Is the system visually appealing? 6) Is the system easy to navigate? There were also two open-ended questions that ask the participant to share any other feedback and to describe problems they might have encountered. Since Task 3 was simply an electronic version of the System Usability Scale, its post-task evaluation omitted questions that were not relevant to this section.

3. Results

Overall, 17 healthy adults at our clinics participated in this study at the end of Q3 2019. Average age of the participants was 33 years old ranging from 22 to 84 years; 12 participants identified as female, 4 identified as male, and one did not provide a response for this question. Participants reported a range of familiarity with informed research consent, from completely naïve users to experienced research coordinators. The average REALM score indicated a ‘High School’ level of health literacy.

Task self-assessment results are presented in Table 1 as average scores, with a score of 5 indicating highest satisfaction. For tasks 1 and 2, content difficulty ranged between 4.4 and 4.7, while question difficulty ranged from 4.7 to 4.8; task satisfaction varied from 4.2 to 4.3; ease-of-navigation ranged from 4.6 to 4.7. One participant required help to complete each of the tasks.

The heuristic evaluation’s highest mean scores were 4.7 for Visibility and 4.8 for Matching between the system and real world, whereas the lowest mean scores of 4.1 and 4.2 were associated with Help Documentation and Controls, respectively.

Table 1. Task Self-Assessment	Mean (SD)
Task 1: Complete Education Module	
Content Difficulty	4.7 (0.7)
Questions Difficulty	4.8 (0.7)
Satisfaction	4.2 (1.0)
Amount of Time	4.1 (1.2)
Visually Appealing	4.1 (1.0)
Easy to Navigate	4.6 (0.8)
Task 2: Complete Consent Module	
Content Difficulty	4.4 (0.9)
Questions Difficulty	4.7 (0.8)
Satisfaction	4.3 (1.1)
Amount of Time	3.9 (1.3)
Visually Appealing	4.4 (1.1)
Easy to Navigate	4.7 (0.5)
Task 3: Complete E-Questionnaire	
Satisfaction	4.8 (0.6)
Amount of Time	4.2 (1.2)
Visually Appealing	4.6 (0.7)
Easy to Navigate	4.9 (0.3)

System Usability Scores (SUS) were normalized in the usual fashion; SUS mean was 84.4, which corresponds with an ‘above average’ usability rating and places this system at greater than the 90th percentile.

The semi-structured qualitative interview captured feedback that described electronic consent on this platform as easy-to-use, engaging, and faster than paper.

4. Discussion

This mobile implementation of the E-Consent platform was well-received by users. The results of the cognitive walkthroughs are largely consistent with our previous work on this platform as well. Indeed, cognitive walkthroughs are well-suited for these evaluations as they effectively focus on how well a naïve / first-time user can perform a task without formal training [6].

The semi-structured qualitative interview yielded numerous actionable points and areas for improvement (Table 2).

The majority of participants reported that the interactive nature of the system improved their attentiveness to the content: “The interactivity helped me pay more attention to it”. Multimedia was also cited as a preferred feature. In the words of one participant, “I really like the videos, I prefer someone speaking to me rather than reading.” And another participant commented “Much better than reading it on paper. Definitely better to have this faster process. More concise is better.”

Although there was wide consensus that this interactive platform is preferable to paper, there was also agreement that human interaction should be maintained in some way: “Humans can build rapport”, “Human interaction is better”.

Regardless, users felt generally comfortable using the system on their own. One participant described an advantage of the interactive platform is that users “...can take their time...” Another described the milieu as “Very comfortable, simple, my parents who are not native English speakers could use it”.

Table 2. Semi-Structured Qualitative Interview, Summarized

Barriers	Advantages	Suggestions
Controls not visible enough	Simple to understand	Instructional video at start
Content too long	Easy to use	Progress bar
Too wordy	Interactive	Additional navigation buttons
Videos too long	Engaging	Subtitles for videos
Videos not descriptive enough	Move at own pace	Pop-up explanation over words
Need more human interaction	Faster than paper	Controls to adjust font size, color
Need paper supplement	Less waste	Streamline, reduce amount of content
Poor sound effect quality	Option to email	Audiobook feature / read-aloud text
Font size too small	Concise	Emphasize compensation of study
Images too small	Easy to navigate	Emphasize risks of study
		Option to mute audio
		‘Save and continue’ feature
		Real-time chat with human

Some participants were concerned about the intuitiveness of the interface for older patients. “Someone older or not familiar with tablets would have problems understanding and using it”. These concerns are reflected in the heuristics evaluations, which demonstrated lowest scores for ‘Help and Documentation’ and ‘Controls’. These scores may be due partly to the smaller display sizes on mobile devices; users may need to scroll the pages to find the buttons to navigate. This also indicates the need for a more obvious ‘help’ feature with instructions that describe this scenario.

Table 3. Heuristics Evaluation Means	
Heuristic	Average Score (SD)
Visibility	4.7 (0.6)
Match (system to real world)	4.8 (0.3)
Control	4.2 (1.3)
Consistency	4.6 (0.6)
Error Prevention	4.6 (0.6)
Recognition	4.6 (1.0)
Flexibility	4.5 (0.7)
Aesthetics	4.6 (0.9)
Error Recovery	4.6 (0.6)
Help & Documentation	4.1 (1.0)

Limitations of this study include the brisk sample size (17 patients) and the relatively high English literacy of those patients as evaluated by the REALM assessment. The mean REALM score was at the ‘High School’ level of literacy, which corresponds with participants’ English Proficiency; all users described their English proficiency as either ‘Excellent’ or ‘Good’. This limitation can be mitigated by broadening the patient base and expanding the study to more locations.

Next steps should be twofold: 1) further refine the interface for mobile devices based on feedback from this study, and 2) expand usability analysis and testing to include more diverse locations with a larger sample size. Electronic consent systems can conceivably reach into patient homes, and so the home setting is another logical next step for evaluation.

5. Conclusion

Formal usability inspection methods continue to generate actionable feedback, which has shaped the E-Consent platform into a highly usable framework for capturing electronic consents and engaging patients. The mobile iteration of this platform has now demonstrated a high usability rating with excellent levels of user-reported satisfaction. To enhance user-friendly design and engage as many patients as possible, future iterations of the product should account for mobile devices.

References

[1] Finkelstein J, Robins D, Liu J. Usability Inspection of Multipurpose Scalable Informed Consent Platform. *Stud Health Technol Inform* **262** (2019), 198-201.

[2] R.E. Mayer. *Multimedia Learning*. Cambridge University Press, New York, 2001.

[3] Bedra M, Wick E, Brotman D, et al. Avatar-based interactive ileostomy education in hospitalized patients. *Stud Health Technol Inform* **190** (2013), 83-5.

[4] Finkelstein J, Cha EM. Using a Mobile App to Promote Smoking Cessation in Hospitalized Patients. *JMIR Mhealth Uhealth* **4(2)** (2016), e59.

[5] Kukafka R, Jeong Ic, Finkelstein J. Optimizing Decision Support for Tailored Health Behavior Change Applications. *Stud Health Technol Inform* **216** (2015), 108-12.

[6] Wharton C, Rieman J, Lewis C, et al. The cognitive walkthrough method: A practitioner's guide. In: Nielsen J and Mack RL, editors. Usability inspection methods. New York: John Wiley & Sons, Inc; 1994. p. 105-40

Towards a Lean Process for Patient Journey Mapping - A Case Study in a Large Academic Setting

Marieke SIJM-EEKEN^{a,1}, Jaqueline ZHENG^a and Linda PEUTE^a

^aAmsterdam UMC, University of Amsterdam, department of Medical Informatics, Amsterdam Public Health Research Institute, Meibergdreef 9, Amsterdam, Netherlands

Abstract. Hospital quality improvement initiatives that focus on the patient experience are increasingly gaining attention. One of the known improvement methodologies is patient journey mapping. Central to this methodology is the mapping process, resulting in a patient journey map: a visual presentation of the complete route a patient follows during all stages of a care trajectory and the patients emotional experience during this journey: the patient journey map. Currently, there is no standard approach on how to perform the steps in the mapping process and evidence is lacking on the most effective and efficient way. As a result, the adoption of the methodology by healthcare professionals is hindered, more time than necessary is spent on the mapping and a lot of significant variation exists in the quality of the resulting patient journey maps. The Lean methodology has a proven track of record to optimize and standardize processes. Hence, by performing a Lean analysis onto the patient journey mapping process, opportunities to optimize and standardize the process may be identified. This study aimed to assess the value of Lean methodology to optimize patient journey mapping in a hospital setting. Observations and interviews were performed to identify the process for patient journey mapping in a Dutch academic hospital. Lean was used to visualize the process and identify inefficiencies. An expert panel reviewed the improvement suggestions. We conclude that Lean can provide value for aligning different views on the process, for structuring the process steps and phases and for identifying optimization actions in the mapping process.

Keywords. Lean, patient journey, quality improvement, patient experience

1. Introduction

Understanding and improving how patients experience their care is increasingly recognized as one of the key components to successfully delivering high quality healthcare. *Patient journey mapping* is a relatively new and upcoming methodology used in hospital quality of care improvement initiatives to identify actions which could improve patient experience with the healthcare system [1]. In mapping a patient journey the experiences with and interaction between the patient and the healthcare providers in all stages of a care trajectory is explored. One of the foundations of this methodology is

¹ Corresponding Author: Amsterdam UMC, University of Amsterdam, department of Medical Informatics, Amsterdam Public Health Research Institute, Meibergdreef 9, Amsterdam, The Netherlands; E-mail: m.e.sijm@amsterdamumc.nl.

that patient experience is not limited to a visit to a specific healthcare provider. It includes the period before and after the visit, such as a visit to a GP, making an appointment for a hospital visit, waiting for the appointment to take place or looking up information in an (online) patient folder. How patients experience their care, is likely to change during this journey. Another central idea therefore is that patient experience is more than the rating that a patient gives to a healthcare provider, it also includes the feelings a patient experiences, like anxiety or uncertainty. The final result of the *patient journey mapping* is a visual representation of the patient journey: a map showing a timeline indicating all relevant interactions between the patient(group) and its healthcare providers (including interaction with their Electronic Systems) in combination with the different aspects of the patients' experience. The review of the patient journey by healthcare professionals subsequently leads to new insights on how healthcare services can better fulfill patient needs. This results in the identification of improvement actions [1,2].

Central to the methodology of *patient journey mapping* is the mapping process. Mapping a patient journey is a highly complex and time-intensive process that consists of multiple stages with interrelated activities to complete [2]. Currently, there is no standard approach on how to perform the steps in the mapping process and evidence is lacking on the most effective and efficient way. As a result, the adoption of the methodology by healthcare professionals is hindered, more time than necessary is spent on the mapping and a lot of significant variation exists in the quality of the resulting patient journey maps [3].

The Lean methodology has a proven track of record to optimize and standardize processes. Hence, by performing a Lean analysis onto the patient journey mapping process, opportunities to optimize and standardize the process may be identified. Because of the shortage in healthcare staff and financial limitations, it is important to work as efficiently and effectively as possible [4,5]. The objective of this study was to assess the value of lean in optimizing and standardizing the patient journey mapping process in an academic hospital setting.

2. Methods

First, the patient journey mapping process was observed and described in a Dutch academic hospital setting. This hospital was selected because of the availability of resources to participate in this research and its recent efforts to define a hospital-wide approach for mapping a patient journey. The scope of this study encompassed all activities from the moment a hospital department initiated a patient journey project till the delivery and review of a patient journey map. The observation was done by an independent Medical Informatician. A Lean tool for visualizing a process, a Swimlane diagram, was used to describe the mapping process. A Swimlane diagram documents process flows with different tasks per role and to provides insight into the order in which tasks are performed [6,7].

In the second step an expert panel consisting of a representative from the Communications department (who developed the approach and toolkit), an Operational Lead, a Medical Lead, a Lean expert (Master Black Belt) and a Medical Informatics junior researcher redesigned the mapping process based on the documented process in step one. In this step a Lean waste analysis was performed by a Lean expert. A Lean waste analysis is a tool which identifies suboptimal activities aiming at streamlining a process [8]. This step resulted in a proposal to optimize the process for mapping the

patient journey based on the applied Lean analyses. The value of applying Lean was then determined by the expert panel review on basis of accepted optimizations and insights gained the process. A Lean improvement was accepted by the expert panel if they considered it feasibility for implementation and if it fit the hospital's vision on patient journey mapping.

3. Results

The observations and interviews provided insight in the process for patient journey mapping in the academic hospital. The communications department took ownership on the mapping process and educated the departments that were supposed to apply it. In this paper this department is referred to as the Quality Improvement (QI) team. There was no top leadership pressure to create patient journey maps. The mapping process consisted of 2 workshops (a kickoff and a follow up workshop), actions to be taken by either the medical lead or operational lead and a toolkit to support the activities to follow. Tasks between the medical lead and operational lead were different depending on the way the specialty was organized. Use of the Swimlane diagram resulted in the identification of three phases with interrelated activities: the initiation, the data collection and the mapping phase.

Overall the Lean waste analysis provided sight on 11 improvements. The expert panel review accepted eight improvements and rejected three. Figure 1 shows the resulting Swimlane diagram. The rectangular boxes in Figure 1 represent activities. Each activity has detailed work instructions.

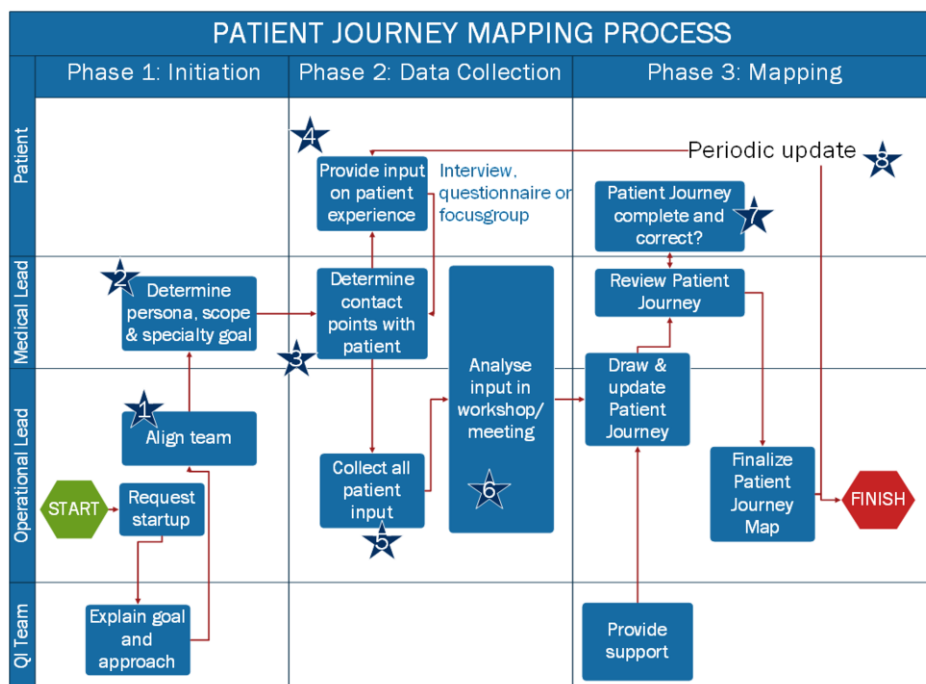


Figure 1. Swimlane diagram showing Lean improvements for patient journey mapping. Stars indicate improvement suggestions, QI Team = Quality Improvement Team

The numbered stars highlight the accepted Lean optimizations. Below each optimization is described in detail.

- *Star 1- Align team upfront*
Prevent wasting time and effort on discussion and confusion about who is supposed to be involved in which actions by agreeing up front on the roles and responsibilities between QI Team, Medical Lead and Operational Lead.
- *Star 2 – Faster determination of persona, scope and goals*
Reduce time and effort spent by healthcare professionals on determination of persona, scope and goals by providing guidance on how to do this.
- *Star 3 – Use available resources to determine contact points with patient*
Prevent having to change the patient journey later on by using available system data to determine contact points with patients.
- *Star 4 – Collect high quality input on patient experience*
Increase quality of data on patient experience by providing guidance on the selection of patients and by standardizing questionnaires for collecting data about patient experience.
- *Star 5 - Only start a workshop when all input is available*
Make sure all required data which is needed in a workshop is available, so that the workshop can deliver expected outputs without wasting time and energy of attendees.
- *Star 6 - Reduce amount of workshops*
Do not organize workshops if the goal can be realized by using other means of communication so that timelines are reduced.
- *Star 7 - Review patient journey map with patient*
Increase quality of the patient journey map by reviewing the map by a patient (or representative) before finalizing it. This prevents rework later on and prevents healthcare professionals from spending time on analysing incorrect or incomplete patient journeys.
- *Star 8 - Periodic update*
In a changing world a patient journey map needs regular updates to stay up to date. Otherwise time is waste on analyzing incomplete or incorrect patient journeys and opportunities to improve patient experiences are missed out.

Rejected improvements included: the creation of online instruction materials to reduce time needed to attend a workshop, the creation of online instruction materials, to replace focus groups and questionnaires for collecting patient input by phone calls and drawing the patient journey map once instead of twice (during workshop on paper/ board and after workshop in electronic format).

4. Discussion

In this study we assessed the value of Lean for improving the patient journey mapping process in an academic hospital setting. First, the Swimlane diagram helped with aligning different views on the current mapping process and with structuring the process activities. Secondly, an expert panel reviewed eight out of 11 Lean suggestions based upon Lean waste analysis as valuable and fitting their hospital vision on patient journey mapping.

Other studies also confirmed the value of Lean for optimizing and standardizing processes in healthcare settings, resulting in quality improvement. The systematic review performed by Glasgow in 2010 showed ample examples of successful quality improvement initiatives based on Lean, varying from improving laboratory tests to radiology processes [9]. A strength of our study is to base the Lean analysis upon observation of an actual patient mapping process from initiation till finalization by an independent researcher. Other strength include the execution of the Lean analysis by a Lean expert and the involvement of an expert panel in reviewing the Lean improvements.

A limitation of the setting of this study is that all healthcare professionals involved were mostly unexperienced with patient journey mapping which is reflected in the process. For example there is a higher need for clear instructions and examples, teams had to get aligned at the startup of a new patient journey process execution, etc. The improvement findings might therefore differ from a setting where an experienced team is executing the process.

On the basis of this work, we conclude that Lean can provide value for aligning different views on the process, for structuring the process activities and phases and for identifying optimization actions in a patient journey mapping process. The mapping process described by us is a first step in optimizing the process for patient journey mapping. Further validation and adjustments are needed to come to an optimal process. We look forward to further develop the process for patient journey mapping and intend to validate the proposed model in the near future.

References

- [1] T.M. Trebble et al., Process mapping the patient journey: an introduction, *Bmj* **341** (2010), c4078.
- [2] D.I. Ben-Tovim et al., "Patient journeys: the process of clinical redesign. *Medical Journal of Australia* **188.S6** (2008), S14-S17.
- [3] M. Musman et al., *Research project: Patient journey mapping in Nederlandse ziekenhuizen*. Universiteit van Amsterdam (2019).
- [4] G. Hindricks et al., Remote monitoring, healthcare costs, and workload for healthcare professionals, *European heart journal* **40** (2019), 1847-1849.
- [5] A. Montgomery et al. Burnout in healthcare: the case for organisational change, *Bmj* **366** (2019): 14774.
- [6] M. Weske, *Business Process Management*, Springer, Heidelberg, 2019.
- [7] R. Robert, *The basics of process mapping*. Productivity Press, New York, 2011.
- [8] R.D. Snee, Lean Six Sigma—getting better all the time, *International Journal of Lean Six Sigma* **1.1** (2010), 9-29.
- [9] J. Glasgow et al., Guiding inpatient quality improvement: a systematic review of Lean and Six Sigma, *The Joint Commission Journal on Quality and Patient Safety* **36.12** (2010): 533-AP5.

User Experience Theories, Models, and Frameworks: A Focused Review of the Healthcare Literature

Blake LESSELROTH^{a,1}, Helen MONKMAN^b, Kathleen, ADAMS^c, Scott WOOD^c,
Audrey CORBETT^a, Juell HOMCO^a, Elizabeth M. BORYCKI^c, Ross SPIER^c, and
Andre W. KUSHNIRUK^c

^aUniversity of Oklahoma School of Community Medicine, USA

^bSchool of Health Information Sciences, University of Victoria, Canada

^cOffice of Human Factors and Ergonomics, Veterans Health Affairs, USA

Abstract. User experience (UX) theories, models, and frameworks (TMFs) help scope user-centered design activities, aid in the selection of constructs and measures, and contextualize findings within a larger knowledge base. However, the fragmentation of literature across disciplines and the inconsistent use of TMFs makes integrating concepts and selecting UX tools challenging. Therefore, we conducted a focused review of the healthcare literature to identify a succinct list of UX-specific TMFs for academic UX researchers and industry practitioners alike.

Keywords and MeSH terms. Human-Computer Interface, Theoretical Models, Ergonomics, human factors, usability, frameworks, user experience.

1. Introduction and Problem Statement

User experience (UX) encompasses user perceptions and responses resulting from the use or the anticipated use of a product, system, or service [1]. Because UX research in healthcare is a high-stakes endeavor, questions arise concerning methods and instrumentation. For example, “what methods best suit the context?” and “how will we know if a technology is effective?” Practitioners can use theories, models, and frameworks (TMFs) to answer these questions. Using Nielsen’s definitions, we define “theory” as a set of statements that explain observable phenomena, “model” as a descriptive simplification of those phenomena, and framework as the lexicon to organize constructs [2].

Collectively, TMFs scaffold UX work by organizing concepts, aiding in method selection, and contextualizing results [3]. Because it can be challenging to stay abreast of the literature, we offer a focused review of relevant TMFs reported in healthcare. Our goals were to (1) provide a succinct list of UX TMFs used to study health information technology (HIT); (2) describe several general models that can meet a wide array of UX research needs; and (3) identify theory-research and theory-practice gaps and opportunities.

2. Methods

We completed a PubMed search from 1946 through to 2019 using the following medical subject headings (MeSH) terms and keywords: “Human-Computer Interface

¹ Lesselroth, Blake, E-mail: Blake-Lesselroth@ouhsc.edu.

OR Ergonomics OR human factors OR usability OR technology acceptance OR user experience” AND “Models, Theoretical OR Models, Psychological OR Information Theory or Systems Theory OR frameworks”. We restricted the results to manuscripts applying a theory, model, or framework to healthcare and limited output to “human subjects” and “English language”. We excluded articles of methods without TMFs.

3. Results

We identified 4369 article titles, selected 307 for abstract review based upon relevance (i.e., if they described technology development or implementation), and 104 for full text review. We identified 101 unique articles speaking to UX and related disciplines including implementation science, software development, and cognitive psychology². Of those articles retrieved, 36 TMFs expressly addressed user experience, interface design, or interface usability.

Table 1 User experience theories, models, and frameworks in healthcare (n=36)².

	Theory, model, framework	Brief description	Reference
Human-computer interfaces and design			
1	Health Human Computer Interactions Framework	Identifies the major components of human-computer interaction and disambiguates user roles.	Staggers et al., 2001
2	Ecological Interface Design Framework	General approach to interface design to simplify information processing and improve cognition.	Effken, 2002
3	Human Centered Distributed Information Design Model	Analysis of human computer interaction at four levels of abstraction.	Zhang et al., 2002
4	Benefits Evaluation Framework	A framework of three quality, two system usability, and three net benefits dimensions.	Lau et al., 2007
5	Formative Evaluation in Instructional Design	A decision-making template for developing formative evaluation strategy.	Street et al., 2007
6	Elements of User Experience	Stratifies dimensions of usability according to the product timeline or information abstraction.	Crutzen et al., 2009
7	Usability Analysis Framework for Healthcare Information Technology	A usability analysis framework designed for inspection of health information technology that identifies and prioritizes potential errors.	Sarnikar et al., 2009
8	Cognitive-sociotechnical Framework	Organizes usability and work into levels of user interaction from individual to group interactions.	Borycki et al., 2010
9	User Centered Design Development Lifecycle	Organization of UX activities according to software development lifecycle.	Yen et al., 2011
10	Technology User Representativeness Functions framework (TURF)	A comprehensive framework that can support a multi-dimensional usability evaluation of HIT to predict safety and impact.	Zhang et al., 2011
11	Usa-Design Model©	A model consisting of four phases in the software design lifecycle.	Merino et al., 2012
12	Health IT Usability Evaluation Model (Health-ITUEM)	A usability evaluation model designed specifically for evaluating the usability of mobile health technologies.	Brown et al., 2013

² For search method and a complete bibliography please contact the corresponding author.

	Theory, model, framework	Brief description	Reference
13	Senior Technology Acceptance Model (STAM)	Adaptation of the TAM to elderly target users.	Chen et al., 2014
14	Use of Technology for Adaptation by Older Adults and Those with Limited Literacy (USABILITY)	A model based upon the Technology Acceptance Model and the Theory of Planned Behavior to predict technology adoption by elderly patients.	Caboral-Stevens et al., 2015
15	Health Information Technology Frustration Incidence Model	A holistic approach to understanding users' frustration with healthcare interfaces, recognizing that safety and patient outcomes are at stake.	Opoku, 2015
16	Organization for the Review of Care and Health Applications Framework (ORCHA-24)	A framework for evaluating the quality and risk of health applications based upon observed indicators of information security, clinical efficacy, user engagement, and user experience.	Leigh et al., 2017
17	Factors Affecting Safety and Performance in Human-machine Interactions	A framework of activity system components used to inventory and predict factors influencing safety and performance of human-computer interactions.	Stowers et al., 2017
18	Model for consumer use of medical devices	A model that describes four dimensions of a device and a user that promote safe use practices.	Reyes et al., 2018
19	MOLD-US Usability Framework for Elders	Aging barriers framework for usability problems encountered by older users.	Wildenbos et al., 2018
Technology implementation and acceptance			
20	Technology Acceptance Model (TAM)	A model that characterizes usability determinants predicting attitudes towards technology adoption.	Davis, 1989
21	Technology Acceptance Model 2 (TAM2)	An added variables TAM that identifies the system variables mediating technology adoption.	Venkatesh et al., 2000
22	Diffusion of Innovation Theory	A theory intended to explain how, why and at what rate innovations spread.	Rogers, 2003
23	Human, Organization, and Technology-Fit framework (HOT-Fit)	A model that evaluates how individuals, organizations, and technologies interact to influence HIT adoption.	Yusof et al., 2006
24	Contextual Implementation Model	Human-computer interaction model looking at three levels of interaction: user, unit, and system.	Callen et al., 2008
25	Combined Technology Assessment Model and Theory of Planned Behavior	Combines TAM and TBP to produce a complete set of determinants influencing technology use.	Duyck et al., 2008
26	Normalization Process Theory	Identifies, characterizes, and explains mechanisms that promote or inhibit implementation.	May et al, 2009
27	Framework for testing electronic adherence monitoring devices	A conceptual human factors framework that incorporates objective and subjective dimensions of electronic adherence monitoring technology.	DeBleser et al., 2011
28	Capability, opportunity, motivation – behavior system (COM-B)	A behavioral-change framework that sits at the ‘hub’ of a behavior change wheel and that seeks to overcome limitations of prior frameworks.	Michie et al., 2011
29	Effective Technology Use (ETU) model	An actionable model of technology adoption that identifies five determinants of use behavior.	Holahan et al., 2015
30	Consumer Health Information System Adoption Model	Predicts the influence of user characteristics and perceptions of an innovation on informatics adoption.	Monkman et al., 2015
31	Unified Theory of Acceptance and Use of Technology (UTAUT)	Explains the internal and external factors including users' intention to use an information system.	Venkatesh et al., 2016
Computer science and software engineering			

	Theory, model, framework	Brief description	Reference
32	Model of PC Utilization	Identifies factors influencing computer utilization including complexity and user affect.	Duyck, 2008
33	Information System Success Model	Identifies the processes and attributes of represented information mediating system success.	Golob et al., 2008
Systems engineering			
34	Fit Between Individuals, Task, and Technology (FITT)	A framework looking at the interaction between system variables and human-computer interactions to optimize technology implementation.	Ammenwerth et al., 2006
35	Systems Engineering Initiative for Patient Safety (SEIPS)	An extension of the Structure-Process-Outcome framework for understanding the interactions between people and systems.	Carayon et al., 2006
36	Sociotechnical Model for Health Technology	An eight-dimensional model addressing challenges in design and implementation.	Sittig et al., 2011

4. Discussion

Principle Findings and Three Key Models

We identified 36 unique TMFs describing UX studies of HIT. Many articles used TMFs from related fields or pragmatically borrowed constructs without articulating a link to UX theory. We purposefully selected three examples of UX models for further discussion: the User Centered Design Development Lifecycle, the Cognitive Socio-technical Framework, and the Task, Users, Representations, and Functions framework [3-5]. These are “general purpose” models (i.e., provide a range of constructs and explanatory interactions) developed expressly to facilitate UX work at any stage of development. Other models may be better tailored to a particular task or context such as app development or designing for the elderly.

The User Centered Design Development Lifecycle, described by Nielsen, is a cycle of activities fusing iterative design-thinking with the software development lifecycle [6-8]. Yen adapted this model to healthcare by tracing development stages to evaluation methods and healthcare goals [3]. The model describes three phases. The pre-design phase calls for UX instrument preparation during project planning. The design phase includes four steps: (1) specifying user needs; (2) defining requirements; (3) designing components and; (4) combining components. The post-design phase includes two steps: (1) system integration; and (2) routine use (i.e., UX to optimize regular use) [4, 5, 7].

Borycki and Kushniruk’s Cognitive Socio-technical Framework (CSF) integrates two complementary characterizations of user interactions: (1) cognitive models describing how humans process information and; (2) sociotechnical models describing how technology affects work [4]. The framework suggests UX evaluation methods throughout the technology lifecycle. Initially, researchers should study the interaction between technology and individual. Later, researchers can examine work processes among groups in context.

The Task, Users, Representations, and Functions (TURF) framework developed by Jiajie Zhang is a four-step sociotechnical model that emphasizes the interactions between users, tasks, and functions [5]. The first step requires identification of all system users. In the second step, researchers gain an understanding of the work by developing an ontology of activities, interactions, and knowledge. Then, teams can look at software through multiple lenses to identify unmet needs. The third step is a

task analysis wherein researchers catalog tasks associated with complex activities to measure complexity and time on task. Finally, the team can study how effectively data displays represent concepts.

Limitations and Opportunities for Future Research

We only searched PubMed using MeSH terms and keywords; we did not search other bibliographic databases. A future systematic review should include other bibliographic databases (e.g., Scopus, Embase), trade publications, and grey literature.

Future research needs to focus on: (1) schematizing the dimensions of usability; (2) predicting the effect of constructs upon outcomes; and (3) publishing practical guides linking TMFs to use-cases, technology phases, and study methods [9, 10].

5. Conclusions

Application of UX TMFs can help scope projects, address usability issues, and improve overall system performance. However, the state of the literature makes it difficult to integrate concepts and apply to practical problems. Although we provided a list of generalizable models, we have identified opportunities to close the theory-practice gap.

References

- [1] International Standardization Organization. 9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO) Switzerland. 2009.
- [2] Nilsen P. Making sense of implementation theories, models and frameworks. *Implementation science*. 2015;10(1):53.
- [3] Yen P-Y, Bakken S. Review of health information technology usability study methodologies. *Journal of the American Medical Informatics Association*. 2011;19(3):413-22.
- [4] Borycki E, Kushniruk A. Towards an integrative cognitive-socio-technical approach in health informatics: analyzing technology-induced error involving health information systems to improve patient safety. *The open medical informatics journal*. 2010;4:181.
- [5] Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. *Journal of biomedical informatics*. 2011;44(6):1056-67.
- [6] Jokela T, Iivari N, Matero J, Karukka M, editors. The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. *Proceedings of the Latin American conference on Human-computer interaction*; 2003: ACM.
- [7] Hartson R, Pyla PS. *The UX Book: Process and guidelines for ensuring a quality user experience*: Elsevier; 2012.
- [8] Nielsen J. *Usability engineering*: Elsevier; 1994.
- [9] Chung AZ, Shorrock ST. The research-practice relationship in ergonomics and human factors—surveying and bridging the gap. *Ergonomics*. 2011;54(5):413-29.
- [10] Chung AZ, Williamson A. Theory versus practice in the human factors and ergonomics discipline: Trends in journal publications from 1960 to 2010. *Applied ergonomics*. 2018;66:41-51.

User-Centred Approach to Design an Online Social Support Platform for Seniors : Identification of Users' Types and Their Requirements

Arnaud RICCI^{a,1}, Jessica ROCHAT^b, Henk Herman NAP^c, Lotte CORNELISSE^c,
Christian LOVIS^{a,b}, Frédéric EHRLER^a

^a*Division of medical information sciences, University Hospitals of Geneva, Switzerland*

^b*Faculty of medicine, University of Geneva, Switzerland*

^c*Innovation & Technology of eHealth, Dutch Expertise Centre on Long-Term Care, Vilans, Utrecht, The Netherlands*

Abstract. A well-designed social platform has the potential to reduce senior isolation and promote active ageing. However, to design tools that respond to the user need it is important to understand what types of seniors will use it and what are their needs. Through a user-oriented approach at several design stage: focus group, co-creation and usability test we were able to identify the different roles related to the use of the platform, which allowed us to classify them by type, each type having very specific needs. Among the types of users identified, “Active users” are looking for an efficient platform. “Socialiser users” are more passive and mainly interested to participate in activities and are sensitive to trust. “Low skilled users” have limited digital skills and must be accompanied to start with the platform. Finally, “Sporadic users” lack of time to actively use the platform but would use a platform involving a large number of stakeholders. It is important to include all these different types of users in the design phase to ensure the future success of the social platform.

Keywords. Aged, social participation, social media, needs assessment, user-computer interface

1. Introduction

While good quality of life among seniors is correlated with an active social life [1], loneliness is strongly associated with depression, thus decreasing well-being of elderly [2]. Unfortunately, prevalence of loneliness in Europe is still high since it has been measured at 9.6% in older adults [3]. One way to fight isolation is to strengthen the social link between communities. Researches showed that participation in social activities can be motivated by online social support platforms [4,5]. While some researches have studied seniors' requirements to use social media [6-9], only few studies have focused on the different type of seniors' users and their heterogeneous needs. *PaletteV2* is a

¹Corresponding author: E-mail: Arnaud.Ricci@hcuge.ch

European project aiming to develop *Palette*, an online social support platform for seniors to encourage them to an active and social life. The aim is to provide to seniors a user-friendly online platform that gives them access to likeminded people and activities, aligned with their interests. They can create different types of events such as recreational or volunteer events and search for their activities according to their interests and by type of event. They can also easily discuss with the creator of the activities.

It is known that ageing influences how people feel comfortable using technologies and how they learn to use new technologies [10]. To ensure the success of any product and its adoption by users, it is essential to understand end-users needs and to design and evaluate it with a range of potential end-users at each stage [11]. An iterative user-centered design process took place in which the interface was iteratively ideated, designed and evaluated with end-users at each stage of the project. In this article, our goal is to present the different users types of such platform and their requirements.

2. Methods

The development of the social platform named “*Palette*” was conducted as an iterative process involving targeted users at each stage. First, focus-groups allowed us to identify user’s difficulties and needs [12]. This stage highlighted their desire of seniors to engage in meaningful and social activities. Then, co-creation sessions were organized to understand how an online platform could support this goal and to ideate the platform functionalities with end-users [9]. Finally, prototypes and final platform were iteratively tested with end-users to assess and improve the usability [13].

During this whole process we collected a large amount of information about our user group and aimed to classify them into categories regarding their intended and actual use of the platform. Then we associated these categories with their main need and barrier regarding the platform

Although there is no strong consensus on a single taxonomy, there are several attempts to classify social network users in particular categories depending on their behavior. To classify our users, we were guided by the research of Brandtzaeg & Heim (2011) [14], which define five roles of users on social networks: sporadics (new comers and users who use the platform irregularly), lurkers (people using media for lurking or time-killing), socialiser (early majority, intermittent contributors), debaters (early adopters, intermittent contributors), active (users who actively participate on social networks).

3. Results

3.1. *Participants*

In total, 165 end-users participated in the participatory design process of the platform. Focus groups were held in 3 countries, Switzerland, the Netherlands and Romania from August to December 2016. The first session was organized in the Netherlands and Switzerland and included 4 participants each time (average age = 75 years) and the focus groups that followed included 2-3 sessions each with 4-6 participants aged 50 to 80 years (average age = 64 years). This represents a total of 41 participants. Co-creations sessions took place in March 2017 in the same countries with a total of 47 participants (mean age = 67 years). Concerning usability testing of the platform, it involved 5-10 participants

ages between 50-80 years old (mean age = 68 years) in each country with a total of 77 participants.

3.2. Types of users and their requirements

During the different design phases, we were able to classify our participants into 4 categories guided by the Brandtzaeg & Heim model (2011). However, we had to create a new category to consider users who do not have the necessary skills to use the platform despite their willingness to use it.

Some participants did not have the necessary skills to get involved, others wanted to get involved minimally, others wanted to be actively involved and finally participants did not have the time to get involved. Based on these 4 types of users we associated their need and barrier from the data collected at the different design stages,

Table 1. Type of users, needs and barriers

<i>Type of user</i>	<i>Needs</i>	<i>Barrier</i>
<i>Low skilled user</i>	<i>Improving digital skills</i>	<i>Independence</i>
<i>Socialiser user</i>	<i>Platform offering a wide range of activities and a wide variety of offers</i>	<i>Trust</i>
<i>Active user</i>	<i>Easy-to-use platform with feedback and quickly and easily publish and evaluate events</i>	<i>Poor platform ergonomics</i>
<i>Sporadic user</i>	<i>Have all the activities organized in the municipality</i>	<i>not a lot of time (socially active) and very busy</i>

Low skilled user - seniors with poor digital skills

Due to their poor digital skills, some of these seniors do not have an email address. They also have difficulty using a keyboard or a mouse. People with such profile needs to be helped in order to use the platform. This category of seniors considers that the use of the platform could be a good way to improve their digital skills. To use the platform by their own, they would need a support functionality. Organizing meetings to teach them the basics of computer use in addition to test the platform could be beneficial to them and could allow this category to benefit from this service. This is a category that is not found in the Brandtzaeg & Heim (2011) model because their model does not take into account people who do not have enough skills to use social networks but would like to use them.

Socialiser user - seniors who just want to participate in activities

This category can be considered as poorly qualified regarding their digital skills. They sometimes have doubts about Internet security. Some of them don't feel safe trying things on the Internet and they are afraid of making mistakes. They are interested in the concept of the platform and want to use it to meet people and find activities but they don't want to organize events themselves. For these seniors, it's necessary to offer a platform offering a wide range of activities and a wide variety of offers, including support services. They would like to have a system that shows all the offers grouped by municipality or region. This category of users' needs to know that the platform is

trustworthy. One way to do this is to use a moderator to moderate the content. Ideally, it should be a local person who can be contacted by phone or email.

Active user - seniors who are active and want to offer events

This profile of seniors is active and wants to help others, invite others to join activities in which they have already participated or to propose activities they organize themselves. They are often more comfortable with digital than the others. They are often effective in explaining to other seniors how the platform works. The platform must be easy to use with system feedbacks. For example, that tells them that they have correctly created their events and to make sure that other users have access to them. For people who offer several activities it's important to quickly and easily publish and evaluate events. In particular, it should be possible to create an event without a specific start and end date.

Sporadic user - seniors who are socially active and too busy

This category of senior are very socially active with a busy schedule, they don't feel the need to use such platform. They don't really have something to offer and they don't need to use it to have social contacts and participate in activities. Although this category of seniors is not the most concerned by the platform, it can be useful for them to have a platform where all activities of municipalities and organizations are centralized and displayed. The platform could be used by some as an online agenda and replace the activities posted on the different websites.

4. Discussion

Seniors attitudes' and uses' of digital tools is known to be heterogeneous [15]. The user-centered approach used throughout the project allowed us to identify four types of users to consider when creating an online social platform for seniors. Each type have specific needs and thus a different use of an online social platform.

Among the types of users identified, the "*Active users*" want to participate actively to the platform by providing a lot of activities. For them, the most important is to have an efficient platform to avoid losing time in using the platform. To satisfy this need, it is important to be able to filter by categories the different events proposed. It is also important to be able to sort according to the user's interests, otherwise the platform becomes complicated to use when it offers many events. "*Socialiser users*" are more passive and mainly interested to participate in activities. They are looking for various content and are sensitive to the trust in the platform. To address trust issue, a moderator role is important to moderate the content and strengthen users' trust in the platform. "*Low skilled users*" have limited digital skills, highlighting the importance of offline training sessions to help to get started with the platform. It is also necessary to ensure that the various functionalities are easy to access and use. Online tutorials should also be offered to these users. Finally, "*Sporadic users*" are already socially active and lack of time to actively use the platform, they would still use it as an event reference if the municipality's activities were listed on it. This requires involving stakeholders like municipalities as event provider in platform if we want to attract this type of senior.

When developing such platform, it is important to include all types of end-users into a user-centered process to make sure the tool meet the needs of all users.

References

- [1] M. Levasseur, L. Richard, L. Gauvin, É. Raymond, Inventory and analysis of definitions of social participation found in the aging literature: Proposed taxonomy of social activities. *Soc Sci Med* 71 (2010), 2141–2149. A.N. Author, Article title, *Journal Title* 66 (1993), 856–890.
- [2] Cacioppo JT, Hughes ME, Waite LJ et al (2006) Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. *Psychol Aging* 21:140–151
- [3] Eurostat, Union européenne, Active ageing and solidarity between generations: A statistical portrait of the European Union. Publications office of the European Union, 2012
- [4] C. Cachadinha, J. A. Costa Branco De Oliveira Pedro, J. Carmo Fialho, (2011). Social participation of community living older persons: Importance, determinants and opportunities. Proceedings of the 6th International Conference on Inclusive Design "The Role of Inclusive Design in Making Social Innovation Happen", (2011).
- [5] Goswami, S., Köbler, F., Leimeister, J. M., & Krcmar, H. (2010, June). Using online social networking to enhance social connectedness and social support for the elderly. Association for Information Systems.
- [6] Gomes, G., Duarte, C., Coelho, J., & Matos, E. (2014). Designing a facebook interface for senior users. *The Scientific World Journal*, 2014..
- [7] Chou, W. H., Lai, Y. T., & Liu, K. H. (2013). User requirements of social media for the elderly: a case study in Taiwan. *Behaviour & Information Technology*, 32(9), 920-937.
- [8] Karahasanović, A., Brandtzaeg, P. B., Heim, J., Lüders, M., Vermeir, L., Pierson, J., ... & Jans, G. (2009). Co-creation and user-generated content—elderly people's user requirements. *Computers in Human Behavior*, 25(3), 655-678.
- [9] Rochat, J., Nap, H. H., Ricci, A., Cornelisse, L., Lukkien, D., Lovis, C., & Ehrler, F. (2018). Designing an Online Social Support Platform Through Co-Creation with Seniors. *Studies in health technology and informatics*, 247, 760-764.
- [10] Barnard, Y., Bradley, M. D., Hodgson, F., & Lloyd, A. D. (2013). Learning to use new technologies by older adults: Perceived difficulties, experimentation behaviour and usability. *Computers in Human Behavior*, 29(4), 1715-1724.
- [11] Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). Key principles for user-centred systems design. *Behaviour and Information Technology*, 22(6), 397-409.
- [12] R. A. Krueger, M. A. Casey , Focus groups: A practical guide for applied research. Sage publications, 2014
- [13] Bastien, J. C. (2010). Usability testing: a review of some methodological and technical aspects of the method. *International journal of medical informatics*, 79(4), e18-e23.
- [14] Brandtzaeg, P. B., & Heim, J. (2011). A typology of social networking sites users. *International Journal of Web Based Communities*, 7(1), 28-51.
- [15] Czaja, S. J., & Lee, C. C. (2017). The potential influence of the internet on the transition to older adulthood. In *New dynamics in old age* (pp. 239-252). Routledge.

This page intentionally left blank

Section 6

Ethics, Legal and Societal Aspects

This page intentionally left blank

Autonomous Systems and Artificial Intelligence in Healthcare Transformation to 5P Medicine – Ethical Challenges

Bernd BLOBEL^{a,b,c,1}, Pekka RUOTSALAINEN^d, Mathias BROCHHAUSEN^e,
Frank OEMIG^f, Gustavo A. URIBE^{g,h,i}

^a Medical Faculty, University of Regensburg, Germany

^b eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany

^c First Medical Faculty, Charles University of Prague, Czech Republic

^d Tampere University, Tampere, Finland

^e College of Medicine, University of Florida, Gainesville, FL, USA

^f Deutsche Telekom Healthcare and Security Solutions GmbH, Bonn, Germany

^g The European Organization for Nuclear Research, Geneva, Switzerland

ⁱ Telematics Engineering Research Group, University of Cauca, Popayán, Colombia

Abstract. The paper introduces a structured approach to transforming healthcare towards personalized, preventive, predictive, participative precision (P5) medicine and the related organizational, methodological and technological requirements. Thereby, the deployment of autonomous systems and artificial intelligence is inevitably. The paper discusses opportunities and challenges of those technologies from a humanistic and ethical perspective. It shortly introduces the essential concepts and principles, and critically discusses some relevant projects. Finally, it offers ways for correctly representing, specifying, implementing and deploying autonomous and intelligent systems under an ethical perspective.

Keywords. pHealth, P5 medicine, autonomous systems, artificial intelligence, ethical principles

1. Introduction

In many countries, specific initiatives and strategic programs are established and continuously updated, aiming at improving care quality, patient safety, and care process efficiency and efficacy, thereby moving from volume to value based care to respond to the challenges health systems face. Those challenges are, e.g., ongoing demographic changes towards aging, multi-diseased societies, the related development of human resources, a health and social services consumerism, medical and biomedical progress, and exploding costs for health-related R&D as well as health services delivery. Organizational, but especially by disruptive methodological and technological paradigm changes enable this move. The paper shortly introduces in those paradigm changes and the accompanying organizational and ethical challenges and proposes

¹ Corresponding Author. Bernd Blobel, PhD, FAACMI, FACHI, FHL7, FEFMI, FIAHSI, Professor, University of Regensburg, Medical Faculty; Regensburg, Germany; Email: bernd.blobel@klinik.uni-regensburg.de

principles and methodologies for mitigating them, thereby critically discussing some ongoing solutions and projects.

2. Methodological, Technological and Organizational Paradigm Changes

Methodologically, phenomenological and evidence-based medicine, both relying on population-based diagnostic models based on retrospective data sets, slowly evolve towards personalized, preventive, predictive and participative precision (5P) medicine. It considers individual health state, conditions and social, environmental, occupational and further contexts at bedside in relation to the community, fully understanding the specific pathology of the health problem [1]. This requires the multidisciplinary approach of systems medicine, deploying the explicit and enhanced knowledge of all stakeholders from the different domains to be involved including the subject of care in the center, replacing the observational and analytical medicine approach [1]. Christensen et al. described healthcare transformation as move from intuitive through empirical to precision medicine [2]. Involved disciplines/domains include medicine and public health, natural sciences, engineering, administration, but also social and legal sciences and the entire systems sciences world (systems medicine, systems biology, systems pathology, etc.). Organizationally, health systems transform from an organization-centric through a cross-organizational, pre-defined and process-controlled to a context-sensitive, individually tailored, highly dynamic, fully distributed personalized care paradigm. The latter is sometimes also called ubiquitous care or care anywhere at any time. With a stronger focus on the information and communication technology (ICT) support, thereby referring to pervasive computing technology, another term frequently used is pervasive care. Both described paradigm changes are supported, impacted or even enabled by related technological paradigm changes. Here have to be mentioned: Mobile, nano-, bio- and molecular technologies; artificial intelligence (AI); robotics; bioinformatics; big data and prescriptive (based on current data) as well as predictive (includes future outcome) analytics; natural language processing (NLP) and understanding (NLU); cloud computing; cognitive computing and social business; but also the Internet of Things (IoT) [3]. A more detailed description of health systems transformation is provided in [1].

We cannot place health and care specialists, educators, lawyers, etc., next to every person to be comprehensively served. Ubiquitous 5P medicine requires the deployment of robotics and artificial intelligence, or more generally autonomous and intelligent systems (AIS), aiming at [3, 4, 5, 6]:

- a) capability and engagement augmentation for care provider and subject of care including education, access to information and services, thereby advancing accuracy, precision, location independency [7];
- b) enabling cooperation;
- c) improved staff and patient experience;
- d) process improvement including clinical workflow and scheduling, but also business efficiency, productivity and cost containment as well as risk analysis;
- e) facilitating faster and more precise decision at administration, direct and indirect caregiver, and patient level including prognosis;
- f) collaborative business intelligence as self-service.

More details on types of artificial intelligence, services and related challenges especially in the health and social care context are provided, e.g., in [8].

Any action and relationship in enlightened democratic societies, but especially the health, care and welfare system have to accommodate legal, moral and ethical principles. In the next chapters, we will consider the different levels of AIS, reference moral and ethical principles, discuss initiatives and projects tackling the ethical challenges, and finally propose a sound approach for addressing this problem.

3. The Ethical Challenge of AIS

Social contracts and law define and enforce behavior for maintaining social order, peace, and justice in society. Ethics provides code and conduct guiding to decide what is good or wrong, and how to act and behave properly, thereby establishing as well as defending rules of morality and frequently going beyond the law [9]. With the evolution of societies including sciences and technologies, different approaches to, or theories on, ethics have been developed in the framework of meta-ethics, normative ethics and applied ethics. Here, Aristotle’s and Plato’s virtue ethics, Kant’s deontological ethics, Mill’s utilitarian ethics, and Rawls’ justice as fairness ethics have to be mentioned [9]. Ethical values are strongly impacted by culture, social norms and geographic locations. Having the evolutionary characteristics of ethics and the terrific social and technological developments in mind, there is no chance for one global comprehensive standard of ethics. Instead, basic social ethical principles such as dignity, freedom, autonomy, privacy equality and solidarity, or the more technological categories like fairness, robustness, explainability, and lineage have been established. For bridging the gaps (at least partially), trust through transparency is discussed as solution [10]. Societal and policy guidelines help to remain human-centric by supporting humanity’s values and those ethical principles [11]. There are many organizations and initiatives proposing ethical frameworks and design methodologies for AIS, such as the EU Council of Europe with its “Guidelines on Artificial Intelligence and Data Protection”, the Future of Life Institute with the “Asilomar AI Principles”, IEEE with “The IEEE Global Initiative on Autonomous Systems”, the U.S. Congress Resolution “Supporting the Development of Guidelines for the Ethical Development of Artificial Intelligence”, the OECD “Principles for AI Research and Development” proposed by the Conference Toward AI Network Society, April 2015, in Japan, The World Economic Forum “Top Ethical Issues in Artificial Intelligence”, and many others. An overview about those initiatives, some content details and references are provided in [8]. Table 1 summarizes the common principles of some of them.

Table 1. Common ethical principles proposed by different organizations

Guideline Originator	Transparency	Accountability	Controllability	Security	Value Orientation Ethics	Privacy	Safety	Risk	User Assistance
OECD	x	x	x	x	x	x	x		x
IEEE	x	x	x		x	x		x	x
Asilomar	x	x	x	x	x	x	x	x	
US Congress	x	x	x	x		x	x		x
World Economic Forum				x	x			x	

4. Representing/Modeling Ethical, Moral and Legal Concepts

Ethics and morality are complex humanistic domains with social, legal, religious and philosophical impact. Their concepts, relations and constraints on them can be represented but not be defined and justified in ICT viewpoints due to the context-free, highly expressive and formal languages deployed, resulting in inconsistent, indefinite and incomplete models. Instead, the concepts, relations and constraints of the domains contributing to the real-world business system must be formally represented and interrelated/harmonized using the ISO 23903 Interoperability and Integration Reference Architecture approach [12]. Its system-oriented, ontology-driven, policy-controlled formal representation of real-world business systems and the related software development process extends ISO/IEC 10746 Open Distributed Processing – Reference Model and turns it into a multi-domain model [13]. Figure 1a presents the AIS use case to be automatically transformed into corresponding ICT solutions.

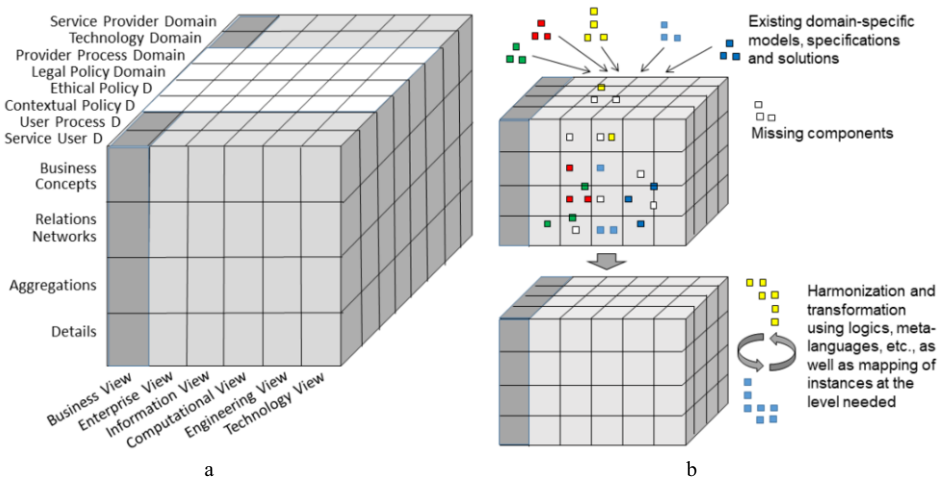


Figure 1. AIS representation, design and implementation enabling advanced interoperability and integration acc. to ISO 23903

Despite the aforementioned limitations regarding the ICT modeling of complex systems, overcome by profiling specifications and remaining at quite generic level, projects have been established to specify implementable components representing ethical and context-related behavior. Here, efforts in modelling morality with prospective logic [14], but also the IEEE P70xx project series [15] have to be named. As they just focus on the ICT representation, they must be integrated and correctly interrelated using the ISO 23903 framework as shown in Figure 1b.

5. Discussion and Conclusions

It is crucial to define objectives, constraints and limits for AIS as well as inevitable principles and not acceptable behavior, auditing their entire lifecycle. In that context, AIS can never be used as means relinquishing or displacing humans' responsibility [10]. The complexity of AIS and the multiplicity of interaction levels for ethical behavior require the consideration of all aspects of human value instead of defining and

implementing a sub-set of legal and ethical principles. Modelling such AIS and their components must be performed following Good Modelling Practices [16, 17], starting with the multi-disciplinary knowledge space, its formalization and harmonization. Many of the ongoing ethics-related projects are limited to ICT perspectives, ontologies [18], and representation styles. The IEEE P7007 project "Ontological Standard for Ethically Driven Robotics and Automation Systems" for example starts with representing some ethical principles and scenarios as UML diagrams, whose elements are then axiomatized using the free and IEEE owned Suggested Upper Merged Ontology (SUMO), developed as foundation ontology for a variety of computer information processing systems. For integration with other specifications, the deployment of the ethical domain ontology represented in the domain's language [19] is inevitable. The current ISO 23903 project provides an appropriate framework also for developing ICT solutions under legal and ethical concerns. A recent formulation of those concerns is provided in [20].

References

- [1] Blobel B. Challenges and Solutions for Designing and Managing pHealth Ecosystems. *Front. Med.* 2019; 6: 83. doi: 10.3389/fmed.2019.00083
- [2] Christensen CM, Grossman JH, Hwang J. *The innovators prescription: a disruptive solution for health care.* New York: McGraw-Hill Education; 2017.
- [3] Weldon D. 10 top analytics and business intelligence trends for 2019. *Health Data Management*, December 07, 2018.
- [4] Siwicki B. IBM Watson Health's chief health officer talks healthcare challenges and AI. *Healthcare IT News*, February 13, 2019.
- [5] Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. April 4, 2019, *N Engl J Med* 2019; 380:1347-1358, DOI: 10.1056/NEJMr1814259
- [6] Davenport TH, Glover WJ. Artificial Intelligence and the Augmentation of Health Care Decision-Making. *NEJM Catalyst* June 19, 2018.
- [7] Cornet J. The Robots in Healthcare Are Here to Stay
- [8] Blobel B, Ruotsalainen P. Healthcare Transformation Towards Personalized Medicine – Chances and Challenges. *Stud Health Technol Inform.* 2019; 261: 3-21
- [9] Tzafestas SG. Ethics and Law in the Internet of Things. *Smart Cities* 2018; 1(1): 98-120.
- [10] Klovig Skelton S. IBM pushes boundaries of AI, but insists companies take an ethical approach. *ComputerWeekly*, 21. November 2019.
- [11] Institute of Electrical and Electronics Engineers (IEEE). *Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition.* IEEE; 2019.
- [12] International Organization for Standardization. *ISO 23903 Health informatics – Interoperability Reference Architecture.* Geneva: ISO.
- [13] International Organization for Standardization. *ISO/IEC 10746-3:2009 Information technology-Open distributed processing-Reference model: Architecture.* ISO, Geneva, 2009.
- [14] Pereira LM, Saptawijaya A. Modelling Morality with Prospective Logic. *Springer's LNCS* 2007;4874:99-111.
- [15] Winfield AFT. Ethical standards in robotics and AI. *Nature Electronics* 2019; 2: 46-48.
- [16] Lankhorst M, et al., *Enterprise Architecture at Work. The Enterprise Engineering Series.* Berlin Heidelberg: Springer- Verlag; 2009.
- [17] Rebstock M, Fengel J, Paulheim H. *Ontologies-Based Business Integration.* Berlin Heidelberg: Springer-Verlag; 2008.
- [18] Akerman A, Tyree J. Using ontology to support development of software architectures. *IBM Systems Journal* 2006; 45 (4): 813–825.
- [19] Arp R, Smith B, Spear AD. *Building Ontologies with Basic Formal Ontology.* Cambridge, Massachusetts – London, England: The MIT Press; 2015.
- [20] High-Level Expert Group on Artificial Intelligence. *Ethics Guidelines for Trustworthy AI.* Brussels: European Commission; 8 April 2019.

Big Data, Information Technology and Information Professionals: Some Considerations for Digital Ethics

E-H. KLUGE,¹

^aIMIA (SiHIS) and University of Victoria

Abstract. It is generally accepted that the global evolution of health information technology, both in design and usage, raises ethical issues that should be addressed. However, the terms in which this concern is expressed are shrouded in ambiguity. Even what constitutes an ethical issue itself is never clearly defined. The present discussion attempts to clarify the landscape and suggests how the concern should be addressed.

Keywords. Information technology, digital ethics, certification

1. Introduction

It is generally accepted that the global evolution of health information technology, both in design and usage, raises ethical issues which should be addressed. Big Data, digital ethics and information technology, as well as the behavior of health informatics professionals and institutions stand at the center of concern. However, as so often happens, the problematic itself as it is usually stated is essentially unclear. For example, what counts as Big Data? What is covered by the phrase “Information Technology?” Who is an information professional? What is meant by “Digital Ethics?” More importantly still, why would anyone think that ethical considerations are worth looking at in this connection? Why not let pragmatic and legal considerations carry the day?

2. Some Clarification

Perhaps the best way to answer these questions is to start with the phrase “Digital Ethics.” It is of course a phrase of art. It refers to the ethical parameters that play a role in the design and use of digital technology. It assumes that there are special ethical considerations that are relevant whenever digital technology is designed or used in informatic contexts in health care.

Like all assumptions, this assumption should be justified. However, doing that assumes that the phrase “ethics” itself is clear and unambiguous—and unfortunately, that is not the case. What people call ethics often has nothing to do with ethics but instead refers to custom, ethos and even to law. Ethics deals with the moral

¹ Corresponding author, University of Victoria, Finnerty Road, Victoria, BC, Canada; E-mail: ekluge@uvic.ca

principles that govern a person's behavior or conduct. This immediately distinguishes ethics from law. While legal considerations may overlap with ethical considerations, and while one may hope that laws, legal traditions, protocols or codes of conduct are grounded in ethical principles, this is not necessarily the case. The reason is that laws, protocols and codes of conduct are rules that are created by a duly constituted authority to regulate the behavior of members of society or of associations. They may be grounded in ethical principles, but that is not necessarily the case. Ultimately they are dependent on the decisions of those who promulgate them. Ethical principles are not thus decision-dependent. They are grounded in what it is to be a person. Therefore they hold irrespective of whether they are recognized in law, tradition or ethos, and independently of whether they are enshrined in legal stipulations, quasi-legal codes or protocols.

There have been various attempts to show how these principles should be reflected in ethical guidelines or codes of practice by providing lists of so-called fundamental values such as privacy [1], justice [2], autonomy [3] etc. There have even been attempts to systematize and collect these values into a matrix applicable to information technology in general and Big Data in particular [4]. However, values are what people consider important. Therefore, they are ultimately psychological in nature. Hence they may differ from individual to individual and from society to society. To have truly objective validity, ethical guidelines and codes should ultimately be grounded in the principles that are enshrined in Universal Declaration of Human Rights.

When these latter principles are particularized to the informatic context, they emerge as Principles of Information Ethics [5]: the Principle of Privacy and Disposition (All persons and group of persons have a fundamental right to privacy, and hence to control over the collection, storage, access, use, communication, manipulation, linkage and disposition of data about themselves.); the Principle of Openness (The collection, storage, access, use, communication, manipulation, linkage and disposition of personal data must be disclosed in an appropriate and timely fashion to the subject or subjects of those data.); the Principle of Security (Data that have been legitimately collected about a person should be protected by all reasonable and appropriate measures against loss, degradation, unauthorized destruction, access, use, manipulation, linkage, modification or communication.); the Principle of Access (The subjects of health records have the right of access to those records and the right to correct them with respect to their accurateness, completeness and relevance.); the Principle of Legitimate Infringement (The fundamental right of privacy and of control over the collection, storage, access, use, manipulation, linkage, communication and disposition of personal data is conditioned only by the legitimate, appropriate and relevant data-needs of a free, responsible and democratic society, and by the equal and competing rights of others.); the Principle of the Least Intrusive Alternative (Any infringement of the privacy rights of individuals, and of their right of control over these data, may only occur in the least intrusive fashion and with a minimum of interference with the rights of the affected parties.); and the Principle of Accountability (Any infringement of the privacy rights of a person, and of the right to control over data about them, must be justified to the latter in good time and in an appropriate fashion.). These Principles govern—more correctly, perhaps, should govern—the development and use of any information technology and the conduct of those involved with it in a professional or institutional capacity.

3. Digital Ethics, Big Data and Electronic Health Records (EHRs)

Which brings the discussion to digital ethics, Big Data and electronic health records (EHRs). Digital ethics is merely information ethics applied to digital data, and Big Data is nothing more than the collection and manipulation of vast amounts of data.

Here it is important to remember something that is often ignored. Technology itself is ethically neutral. It is just that: technology. It is its use that has ethical implications. When it is used in an ethical manner this not only serves a useful purpose but also advances otherwise legitimate aims. For instance, the scientifically valid and ethically appropriate collection, storage, use and manipulation of data can advance the delivery and effectiveness of health care. Likewise, without this modern medical research would essentially be impossible. What is ethically indefensible is collecting and using data sets inappropriately, surreptitiously or without consent—as was done by Cambridge Analytica.

As to why digital ethics takes a special turn in health informatic contexts, the reason lies in the nature of EHRs. Viewed materially, they are merely electronic entities with certain characteristics; viewed informatically, they are the totality of data that have been generated about a given patient: As such, they are patient analogues in data-space; and because of their relationship to their subjects, they are subject to ethical considerations [6].

And this also has practical implications. Because if it had no practical implications, then digital ethics would be a nice conceptual exercise—the sort of thing one would consider in a theoretical university seminar—but it would be pointless to worry about it in real life.

Legal and policy requirements of course also have practical implications. However, they are only as good as their last version, and as the example of the GDPR shows [7] they may be altered over time; and as changes in informatic laws of various countries show, these may even overturned [8-12]. Therefore institutions and professionals who only follow legal or quasi-legal provisions may find themselves in an uncomfortable position when these are changed or amended. And there is also the fact that statutes and similar provisions are meaningless in practical terms until they are interpreted. To be sure, they set relevant frameworks, but what they mean in real terms is a function of how they are interpreted in actual cases. Therefore, as the old saying goes, “To rely on law is to be only as safe as the last case that interpreted the relevant statute.”

This, then, is where ethics is different and where ethics “bakes bread.” An institution that follows the fundamental principles of information ethics in its protocols and provisions will reap three practically important advantages: (1) It will not have to make fundamental changes in its protocols and operational guidelines as time progresses and statutes or quasi-legal provisions change. (2) It will be able to use its protocols in inter-jurisdictional settings because they are not law-based but ethics-based. (3) They may hold themselves out as providing ethically impeccable services—which can be advertised to tremendous commercial advantage. Therefore, again, “Good ethics is good business.” As for professionals, ethically sophisticated professionals make good employees; and it is very difficult to accuse someone—or to find them guilty—of inappropriate conduct when they have acted in an ethically appropriate manner.

4. Conclusion

However, having an ethically valid framework for digital ethics and supplying informatics professionals, institutions and other users with ethical protocols and guidelines is one thing; ensuring that the stipulations that govern the framework are understood and that the guidelines are actually followed is another. That can be achieved only by putting in place an independent and objective body that provides such guidelines and that monitors how all of they functions in actual practice. There currently is no such body. Therefore ethics-in-application essentially remains a haphazard and spotty affair.

However, even if such a body existed, that would only be part of that story. What also is missing is some means of ensuring that professionals and institutions are qualified to follow ethical guidelines and protocols in the first place, and are actually able to operationalize ethically its tenets when using the digital technology as it currently exists or as it changes with the advent new technology and with changes in the structure of health care deliver itself.

Fortunately, some of these considerations have already struck a responsive chord at the global level. The need of providing an overall ethics framework was addressed by the International Medical Informatics Association (IMIA) when it promulgated its Code of Ethics [13], latter which constitutes the overarching framework of the ethics codes of its member associations. As for the need to ensure that professionals and institutions are actually qualified to operationalize this ethical framework, this is what underlies the ethics-certification project that is currently being worked on by a subgroup of IMIA. It is expected that this work will have the endorsement of WHO and associated agencies. Clearly, if and when this it comes to fruition it will have practical implications for Big Data, information technology, information professionals and digital ethics on a global scale. However, whether it comes to fruition only time will tell.

References

- [1] L. Florini. (ed.), *Cambridge Handbook of Information and Computer Ethics*, Cambridge University Press, Cambridge, 2010.
- [2] S. Fleischacker, , *A Short History of Distributive Justice*, Harvard University Press, Cambridge, MA, 2004.
- [3] B. Cburn, *Autonomy and Liberalism*, Routledge , New York 2010.
- [4] A. Zwitter, Big Data Ethics, *Big Data & Society* **1** (2014), 1-6. doi: 10.1177/2053951714559253.
- [5] International Medical Informatics Association, Ethics for Health Informatics Professionals: The IMIA Code, its Meaning and Implications, available at <https://imia-medinfo.org/wp/wp-content/uploads/2015/07/Handbook-for-revised-Code-of-Ethics.pdf>.
- [6] E-H. Kluge, Medical Narratives and Patient Analogues: The Ethical Implications of Electronic Patient Records, *Methods of Information in Medicine* **38** (1999), 1-7.
- [7] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016; available at <https://eur-lex.europa.eu/eli/reg/2016/679/oj> .
- [8] India, Right to Information Act, No. 22, 2005.
- [9] Canada, Access to Information Act (R.S., 1985, c. A-1).
- [10] China, Regulations of the People's Republic of China on Open Government Information Gov, 2007.
- [11] Israel, Freedom of Information Law, 5758-1998.
- [12] Ukraine, The Law on Information (revised in 2011).
- [13] International Medical Informatics Association, Code of Ethics, available at <https://imia-medinfo.org/wp/wp-content/uploads/2015/07/IMIA-Code-of-Ethics-2016.pdf>.

Citizens' Participation in Health and Scientific Research in Switzerland

Jessica ROCHAT^{a,1}, Christophe GAUDET-BLAUVIGNAC^b, Marzia Del ZOTTO^b, Victor RUIZ GARRETAS^b, Vasiliki FOULI^{a,b}, David ISSOM^b, Caroline SAMER^{a,c}, Samia HURST^a and Christian LOVIS^{a,b}

^a *Faculty of medicine, University of Geneva, Switzerland*

^b *Division of medical information sciences, University Hospitals of Geneva, Switzerland*

^c *Division of clinical pharmacology and toxicology, University Hospitals of Geneva, Switzerland*

Abstract. Understanding motivation and resistance factors affecting citizen participation in health and scientific research allows to find solutions to improve citizen engagement and interest in science. Through a survey, we identified the main factors influencing citizens' participation in scientific research, and their wishes to be more informed. Results show that the respondents' reasons to participate in research were altruistic motivations, in line with other studies carried out in developed countries. The main factor influencing the non-participation is the lack of opportunity, highlighting the importance to better inform citizens about ongoing studies.

Keywords. Community participation, motivation, attitude, research, clinical trial, ethics, global health

1. Introduction

Reasons why citizens participate or not in scientific research have been studied in the fields of sociology and psychology. While research has been conducted on motivational engagement in research in Europe [1-3], Brazil [4] and USA [5], no studies have yet investigated the context of Switzerland.

Switzerland is a country with a specific legal context. As not being part of the European Union, it does not have to comply with the same legal framework. However, following the General Data Protection Regulation (GDPR) in Europe [6], a Swiss Federal Act on Data Protection has been created. The relation of the Swiss citizens to their personal data might be country specific and it may have an impact on their willingness to participate in scientific studies.

In this paper, we describe a survey that was developed to understand the motivation and resistance factors affecting citizen participation in health and scientific research but also their wish to be more informed. It aims at understanding the processes behind these decisions and to gain an insight on how to improve this situation.

¹ Corresponding author: Faculty of Medicine, University of Geneva, Switzerland; E-mail: jessica.rochat@hcuge.ch

2. Methods

We developed a survey with questions related to motivational and resistance factors that may affect citizen participation in health and scientific research. Questions and multiple-choice answers were derived from a literature review to identify factors that motivate or dissuade people to participate in scientific studies and clinical trials. The survey was distributed in French to citizens during two general public events held between July and November 2018 in Geneva (Switzerland), aimed at raising citizens' awareness about science.

The survey was separated into three parts: 1) What are their motivations to participate in scientific studies (e.g. surveys, psychology's experiments) and/or clinical trials? These questions were multiple choices with several possible answers. 2) Why did they never participate in scientific studies and/or clinical trials? These questions were also multiple choices with several possible answers. 3) What do they want to be informed about?

Concerning parts 1 and 2, since several answers were possible, the number of responses to each question was divided by the total number of responses to all the proposed answers. This allowed us to compare response frequencies out of a total of 100%.

3. Results

A total of 124 participants completed the survey. 52% were women (n=64), 46% were men (n=57) and 2% (n=3) did not answer to this question. The mean age was 43 years old. The majority of participants were Swiss (66%) and French (18%). Concerning the level of education, 1% of respondents (n=1) did compulsory school, 11% (n=14) went to high school, 72% (n=89) have a higher or university level, and 16% (n=20) did not answer.

58% (n=72) of participants reported they had already participated in a scientific study (e.g. questionnaires, psychology's experiment) while the remaining 42% (n=52) had never participated in a similar study. Among the participants who already participated in a scientific study, 60% (n=43) had participated in questionnaires, 63% (n=45) in an experiment and 19% (n=14) in another type of survey (e.g. interviews). Concerning clinical trials, 9% (n=11) of the respondents stated they had participated in a clinical trial.

3.1. *Motivations to participate in scientific studies and clinical trials*

Scientific studies

They were 72 responders to the question about the motivations to participate in scientific studies. Figure 1 shows that the motivational factors most often chosen were "to help the science" (30%), followed by "to help the organizer of the study" (25%) and "to help people concerned by purpose of the study" (21%). "Financial compensation" represented only 11% of the responses, followed by "to help relatives concerned by the purpose of the study" (5%), "I'm directly concerned by the purpose of the study" (4%) and others (4%).

Clinical trials

They were 11 responders and figure 1 shows that the main motivational factors concerning the participation in clinical trials were “to help the science” (28%), followed by “to help people concerned by the purpose of the study” (24%) and “to help the organizer of the study” (24%). “Financial compensation” and “I’m directly concerned by the purpose of the study” represented only 10% of the responses, followed by “to help relatives concerned by the purpose of the study” (4%).

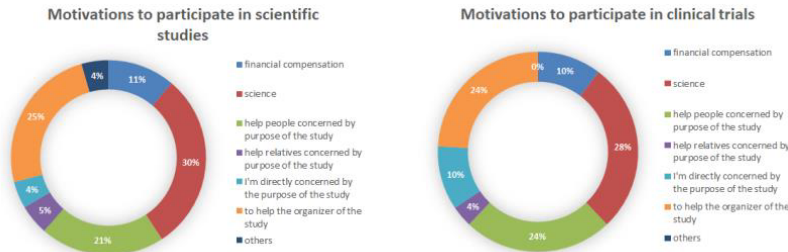


Figure 1. Distribution of motivations to participate in scientific studies (left) and clinical trials (right)

3.2. Resistance to participate in scientific studies and clinical trials

Scientific studies

The total number of respondents to this question were n=52. Figure 2 shows that the main resistance factor influencing the non-participation in scientific studies was the lack of opportunity (86%), followed by the lack of time (6%) and “the fear of what could happen to me” (4%). “The fear that confidentiality and anonymization of my data are not respected” and “the lack of interest for the subject” both represent only 2% of the responses.

Clinical trials

The total number of respondents to this question were n=111. Figure 2 shows that the main resistance factor influencing the non-participation in clinical trials was also the lack of opportunity (69%), followed by “the fear of what could happen to me” (19%). The lack of time represented 6% of the answer followed by “The fear that confidentiality and anonymization of my data are not respected” (5%). The lack of interest for clinical trials represented 1% of answers.

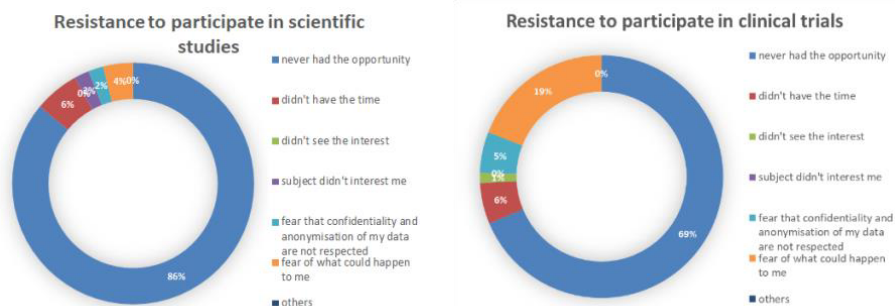


Figure 2. Resistance to participate in scientific studies (left) and clinical trials (right)

3.3. Willingness to be informed about scientific and health research

78% of the participants expressed their willingness to be informed about scientific studies going on in Switzerland. Concerning clinical trials, 70% of participants expressed their willingness to be more informed. Regarding the willingness to be informed about the results of scientific studies, results are quite similar with 77% of positive answers. Concerning results of clinical trials, 71% of respondents answered positively. Regarding more information about ethical committee, 58% of respondents answered positively. 61% of respondents expressed their willingness to be more informed about data privacy management. Finally, regarding more information about data anonymization management, 60% answered positively.

4. Discussion and conclusion

The most frequent motivational factors to participate in scientific studies and clinical trials could be defined as altruistic motivations (e.g. help the science, the people concerned by the purpose of the study and the organizer of the study). These findings are in line with other studies demonstrating that altruistic motivations are the most important motivational factors in developed countries [1, 2, 5].

Individualistic motivations were rarely mentioned in our study both for scientific research and clinical trials, in contrary to other studies where financial compensation was the main motivational factor to participate in health research [7]. According to studies, low educational level and low socio-economic status and young age appear to have a positive link with financial incentives [7, 8]. Our finding can be explained by the high educational level (72%) of our respondents. The main factor influencing the non-engagement on scientific research or clinical trials was the lack of opportunity. This finding is in line with other researches [3, 9] and highlights the need to better inform citizens about ongoing studies. Moreover, the high percentage of respondents that would like to be more informed about ongoing studies, their results, data privacy management, data anonymization management and ethical committee, shows the poor communication between scientific community and citizens. The second factor influencing the choice not to participate in clinical trials was the fear of what could happen (19%), which was much higher than for scientific studies (4%). This result is in line with findings of other studies, underlining that the psychological or physical risks to enroll in clinical trials were one of the major reasons to decline participation [10-11].

It is therefore important to clearly explain the risks and benefits to participants in an understandable language intended for lay people.

Some limitations should be taken into account regarding the results' interpretation. The high frequencies of altruistic motivations compared to individualistic motivations may be partly due to the social desirability bias driving respondents to be viewed favorably [12]. Moreover, respondents were recruited using convenience sampling method (by physical location) during events aimed at raising citizen's awareness about science. This method may impede the generalizability of the results [13]. Also, the number of respondents who participated in clinical trials was low (n=11). Therefore, a broader study needs to be performed.

In conclusion, these results show that to improve citizen participation in scientific studies, the focus should primarily be put on communication to reduce the lack of opportunity. Being able to explain research and science better to lay people as well as the implications of participating in a study is crucial. New initiatives around general and dynamic consent could help to move forward on this topic.

5. Funding

This project is supported by the Swiss National Science Foundations 002632-000-BEHALF; the Swiss Personalized Health Network project 2017DEV 15 – C3Study and the Leenaards foundation project “smartConsent”.

References

- [1] M. Dixon-Woods, C. Tarrant. Why do people cooperate with medical research? Findings from three studies. *Social science & medicine* **68** (2009), 2215-2222.
- [2] L. Baker, T. Lavender, D. Tincello. Factors that influence women's decisions about whether to participate in research: an exploratory study. *Birth* **32** (2005), 60-66.
- [3] L. Almeida, B. Azevedo, T. Nunes, M. Vaz-da-Silva, P. Soares-da-Silva. Why healthy subjects volunteer for phase I studies and how they perceive their participation?. *European journal of clinical pharmacology* **63** (2007), 1085-1094.
- [4] S. A. Nappo, G. B. Iafrate, Z. M. Sanchez. Motives for participating in a clinical research trial: a pilot study in Brazil. *BMC Public Health* **13** (2013), 19.
- [5] M. C. Soule, E. E. Beale, L. Suarez, S. R. Beach, C. A. Mastromauro, C. M. Celano, S. V. Moore BA, J. C. Huffman. Understanding motivations to participate in an observational research study: Why do patients enroll?. *Social work in health care* **55** (2016), 231-246.
- [6] P. Voigt, A. Von dem Bussche. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing* (2017).
- [7] L. Stunkel, C. Grady. More than the money: a review of the literature examining healthy volunteer motivations. *Contemporary clinical trials* **32** (2011), 342-352.
- [8] N. E. Kass, R. Myers, E. J. Fuchs, K. A. Carson, C. Flexner. Balancing justice and autonomy in clinical research with healthy volunteers. *Clinical Pharmacology & Therapeutics* **82** (2007), 219-227.
- [9] K. Lowton. Trials and tribulations: understanding motivations for clinical research participation amongst adults with cystic fibrosis. *Social Science & Medicine* **61** (2005), 1854-1865.
- [10] K. A. Cunny, H. W. Miller. Participation in clinical drug studies: motivations and barriers. *Clinical therapeutics* **16** (1994), 273-82.
- [11] E. Novak, C. E. Seckman, R. D. Stewart. Motivations for volunteering as research subjects. *Journal of Clinical Pharmacology* (1977).
- [12] O. Sjöström, D. Holst. Validity of a questionnaire survey: response patterns in different subgroups and the effect of social desirability. *Acta Odontologica Scandinavica* **60** (2002), 136-140.
- [13] A. S. Acharya, A. Prakash, P. Saxena, A. Nigam. Sampling: Why and how of it. *Indian Journal of Medical Specialties* **4** (2013), 330-333.

Designing a Solution to Manage Electronic Consent for Children

Gary LEEMING,^a Sarah THEW^{a, b} and John AINSWORTH^a

^a*Centre for Health Informatics, Division of Informatics, Imaging and Data Science, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom*

^b*Storybook UX Ltd*

Abstract. Electronic systems for managing consent do exist but are generally only able to record consent from the research subject directly. Consent for research is also challenging to integrate into many electronic patient record systems. The Born In Bradford study is a large, from birth cohort study in the North of England which requires consent to be recorded by the pregnant mother of a child who will be included in the study from birth. This creates a complex challenge for consent management that has previously been achieved through paper-based processes. As the study begins a new phase with the objective of inviting all new parents within the Bradford region to participate in the study the solution also needs to work with existing maternity systems. This paper considers the specific challenges of converting the often grey rules around consent of children into an electronic system that is transparent and supports the trust of both the family and the clinical and care teams recruiting research subjects into a large cohort study, and describes the user centred design and technical approach taken to resolve it.

Keywords. Digital consent, User centred design, UCD, Electronic Patient Records

1. Introduction

The first Born In Bradford study began in 2010¹. Following its success, a second birth cohort study, Born In Bradford 2019² (BiB) began recruitment. BiB will give every mother in Bradford the opportunity to participate in a second cohort which will collect and link patient records from across NHS organisations for both mother and child, and also link to education and other data sources. The BiB team identified that continuing to use paper-based processes for collection and management of maternal consent would be impractical, but had not found a digital solution able to meet their requirements, such as easy integration with their electronic medical record (EMR) system and their existing research dataset. Most systems record consent either directly within the EMR (usually only for medical treatment) or within the study dataset, making it inflexible, hard to maintain and potentially lacking transparency. Consent by mothers for children, especially when planned over the lifetime of the child, is a relationship that must adapt to changes in the child's life. Connected Health Cities³ (CHC) was established to test learning health systems at scale across the North of England, and has worked with BiB 19 to develop a solution to address the complex consent management requirements of a birth cohort study.

2. Methodology

We adopted a user-centred approach to the design and development of the consent system. User centred design (UCD) is a collection of methods emphasising end users of a system and their goals, over the technology, and provide mechanisms for the users to be deeply involved in the design, development and testing of a system. A number of key principles inform this approach⁴:

Early focus on users and tasks: it was important for us to work with users to understand their existing and to-be working practices, environment, tasks and goals.

Empirical measurement: the testing and measurement of current performance, and throughput, e.g. numbers of births, recruitment rates, as well as testing of prototypes.

Iterative design - following cycles of investigation - design - build - measure. This approach fits well with Agile software development practices.

Our starting point was a project charter jointly developed by the teams from BIB and CHC, which set out key functional and nonfunctional requirements, including legislative and regulatory constraints. The project charter was used to derive a programme of UCD work: interviews and workshops with users to explore and document existing working practices, and develop a shared understanding of the requirements implied by the project's Standard Operating Procedures (SOPs.) We co-developed a series of artefacts:

- user stories to understand stakeholders' motivations, goals and context
- process maps to document current and future workflows
- system architecture diagrams, showing where different types of data would be stored
- low fidelity prototypes of user interfaces

These artefacts were used to support discussions with users, and iteratively updated to reflect our emerging joint understanding.

3. Consent Requirements and Implementation

Consent is a key component for people to be able to participate in research, as defined in the Declaration of Helsinki.⁵ BiB is a birth cohort study that does not involve a direct clinical intervention in the consented subject but consent for data access is still required. The guidance for consent of children from the MRC is that a legal guardian can provide consent for themselves and their children.⁶

The process for recruiting subjects into the BIB 2019 study has been to train midwives to discuss the project with all women during pregnancy and, if they consent, to record this within the maternity EMR. Only the mother or a guardian is required to provide this initial consent, which allows for both the mother and child's current and future medical records to be collected, de-identified and used in future research.

Understanding how to codify and implement complex consent requirements is challenging: developers do not have the detailed domain knowledge necessary, whilst the study team are used to working with paper or simple spreadsheets. An iterative UCD approach allowed us to jointly explore and resolve these challenges. Through the design process with stakeholders different requirements were mapped, including:

The data must be kept up to date and accurate. As far as possible the identity details, for example a person's address, are checked against hospital systems connected to the

NHS spine at regular intervals and also when the consent for a specific individual is requested.

The key for identity within the NHS is the NHS number. However, children are not given an NHS Number until they are born and so their consent and identity record needs to be directly linked to their mother and updated in the EMR, and then our consent store at the time of birth, once an NHS number is allocated. This is also linked to a birth order in the case of multiple births.

Children could be born out of area. This means updating of the child’s identity record will not be possible through the local hospital IT systems and must be manually investigated if the maternity record shows a pregnancy longer than 10 months (implying birth may have taken place out of area.)

Consent is given by mother for both herself and the child. Alternatively, a young or vulnerable mother may have a guardian who can give consent for the mother and her child. This means that both consents need to be managed, and the relationship between the consent owner and the child recorded. When withdrawing consent a later requirement was added to ensure that the person requesting the change and their relationship to the data subject is also recorded.

Mechanisms to request and integrate data from external agencies (e.g. education) must minimise the information shared externally.

Two types of withdrawal of consent that are defined. Firstly, a subject can ask to stop future data collection, but agree that existing data will remain within the cohort dataset. Secondly consent can be completely withdrawn meaning the collection of data is halted and the records associated with that consent are deleted.

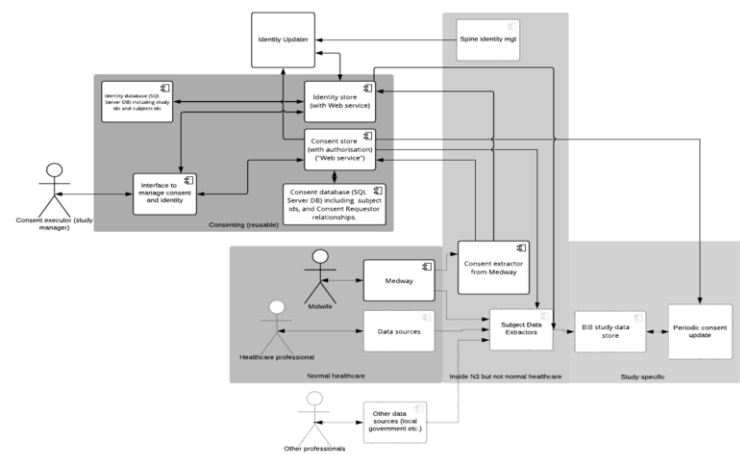


Figure 1. Consent Process Technical Scope including existing NHS systems, BIB 19 consent management and pseudonymised research datastores.

Other considerations include the death of a participating mother or child, and monitoring other changes in the child subject’s status, such as being taken into care. In this latter case the parents of the child may no longer have the right to give or withdraw consent. This responsibility is now that of the new guardian of the child. There is a further question as to when the child becomes responsible for consent and can be considered competent to manage consent for themselves. The age at which the child becomes solely responsible for their consent is defined as 16 although they do not need to be re-consented, in-line with other approaches and the Gillick test in the UK.^{7,8}

Where a guardian previously gave consent for the mother and her unborn child, when the mother reaches the age of 16 she must be re-consented.

The technical implementation of the consent model was based on two main data stores: Identity and consent in order to reflect the need to manage the relationships and consent through distinct mechanisms, as well as make it easier to logically separate identifiable data from the consent. Figure 1 describes how these data stores are managed and updated through a local administration web application, by the study team, and how it interfaces with numerous NHS systems to manage identity and provide core services for the pseudonymised research data set.

Once consent has been collected this gives permission for data to be collected about both the mother and the child. The processes for this data collection are currently in place within Bradford Teaching Hospitals NHS Foundation Trust and cover a variety of sources, including hospital and primary care medical records. This data is stored separately, in a de-identified format using a key linking to the original consent. The record of consent is regularly checked via an API that was developed to match a subjects' study id, and check any date constraints for consent. Standard Operating Procedures (SOPs) for managing consent have been developed and technical changes to the forms within the maternity EMR were agreed. The key requirement for the consent store is to enable the data management team to regularly check for changes in consent amongst the study cohort data in order to be able to update the research dataset.

Finally, there is a web-based administration application which enables the study administrators to manage subjects' details and consent preferences directly.

4. Conclusions and Challenges

The basic principles of consent are clear, but the numerous edge cases and changes to consent that can occur through a child's life present challenges in the design of a computerised consent solution for birth cohort studies. Furthermore, the translation of soft, paper-based processes to computer-based algorithms requires upfront discussion. Interpretation of rules and regulations that are not well codified is challenging, especially for consent for children, which can also vary between different jurisdictions.⁹ There are related challenges to guardianship, divorce, and changes in legal status. Using a UCD approach allowed the project teams to negotiate a joint understanding of these complex requirements, and provided users and developers with a clearer understanding of the implications of the legal requirements and study procedures on the architecture and software implementation.

Benefits include the ability to audit a complex area of consent more closely and simplified management of a large cohort. The solution integrates with the midwives' existing workflow and EMR systems, simplifying subject recruitment. The solution has been designed to be flexible and meet a wide range of edge cases, including integration with different EPRs and research data systems using APIs, making it flexible for the future. The code is made available under open source licence¹⁰ and continues to be improved based on users' feedback.

Working closely with the software team using UCD ensured that the platform met the requirements of the study team while also consulting other stakeholders, such as information governance, to provide a new type of consent management tool necessary for the scale, self-management and linkage of family and consent given for multiple

individuals. In order for this, or similar consent stores, to be able to deliver even greater value for research there is a need to define consent processes more precisely, especially consent for children, within ethical guidelines and law in order to be able to better codify and audit consent against best practice. This will be increasingly important with regards to international co-operation and standards for use of EHR data in research.

Acknowledgements

Born in Bradford is only possible because of the enthusiasm and commitment of the Children and Parents in BiB. We are grateful to all the participants, health professionals and researchers who have made Born in Bradford happen. The authors are also grateful to the staff at Bradford Teaching Hospitals NHS Foundation Trust and the development team David Kemp and Peter Crowther for their essential input into the project. Connected Health Cities is funded by the UK Department of Health and Social Care.

References

- [1] Raynor P, Group B in BC. Born in Bradford, a cohort study of babies born in Bradford, and their parents: Protocol for the recruitment phase. *BMC Public Health*. 2008;8(1):327. doi:10.1186/1471-2458-8-327
- [2] Lane D. Born in Bradford (BiB): A data linkage cohort study of babies Born in Bradford and their mothers : Information for pregnant women. <https://borninbradford.nhs.uk/wp-content/uploads/A3-Participant-booklet-04.10.18.pdf>. Published 2018. Accessed October 11, 2019.
- [3] Connected Health Cities Programme. <https://www.connectedhealthcities.org>. Accessed October 14, 2019.
- [4] Gould JD, Lewis C. Designing for usability-key principles and what designers think. *Conf Hum Factors Comput Syst - Proc*. 1983;28(3):50-53. doi:10.1145/800045.801579
- [5] Kong H, West S. WMA Declaration of Helsinki- Ethical Principles. *World Med Assoc*. 2013;(June 1964):29-32.
- [6] Medical Research Council. Medical research involving children. *MRC Ethics Guid*. 2004:49.
- [7] Baskaran V, Davis K, Bali RK, Naguib RNG, Wickramasinghe N. Managing information and knowledge within maternity services: Privacy and consent issues. *Informatics Heal Soc Care*. 2013;38(3):196-210. doi:10.3109/17538157.2012.735732
- [8] Wheeler R. Gillick or Fraser? A plea for consistency over competence in children. *BMJ*. 2006;332(7545):807. doi:10.1136/bmj.332.7545.807
- [9] Essén A, Scandurra I, Gerrits R, et al. Patient access to electronic health records: Differences across ten countries. *Heal Policy Technol*. 2018. doi:10.1016/j.hlpt.2017.11.003
- [10] CHC Consent. <https://github.com/connectedhealthcities/consent>. Accessed October 11, 2019.

Effectiveness of Web Accessibility Policy Implementation in Online Healthcare Information

Muzawir ARIEF^{a,1}, Sari RISSANEN^b and Kaija SARANTO^a

^a*University of Eastern Finland, Department of Health and Social Management, Finland*

^b*University of Eastern Finland, Faculty of Social Sciences and Business Studies, Finland*

Abstract. UN have recommended the adoption of Web Content Accessibility Guidelines (WCAG) version 2.0 from W3C to guarantee that web content is more accessible to everyone. This study aims to evaluate the effectiveness of WCAG 2.0 implementation by comparing the country with and without WCAG2.0 policy adoption. The objective is to compare the impact of adopting the WCAG 2.0 standard in health information provision by analyzing the differences between countries that adopted the WCAG 2.0 standard and those that did not, specifically for health information related to the elderly. To this end, searches were performed on the Google search engine for online health websites with the keyword “Alzheimer” and the specific country settings for Indonesia and UK, in the local language of each country. Website evaluations were performed for ten websites found with this search by using the WCAG 2.0 measurement tool Axe. Statistical analysis using descriptive and Mann-Whitney analysis to measure the impact of the WCAG 2.0 showed a predominance of low violation occurrences in the UK with 40% from selected websites compared to 80% medium violation occurrences and no low violation occurrences in Indonesia. Although the country with WCAG 2.0 implementation had a lower frequency of violation, no significant differences were found between countries and the media type in WCAG 2.0 evaluation, implying the need to improve the effectiveness of policy implementation.

Keywords. Web Accessibility, Alzheimer, Policy, Online, Information

1. Introduction

Seeking health information online has become the most popular Internet activity among patients [1] and mobile phone users [2]. Previous studies have shown that the majority of people used commercial websites (71.8%), followed by search engines (11.6%), while only 5.5% use government websites for health information [3]. In addition, users with serious health problems and/or disabilities were more enthusiastic about online health information regarding their own health condition [4][5]. Previous studies have suggested that older people use the Internet the least due to many determinants such as age, physical limitation, education, social economic situation and social situation [6].

¹ Corresponding Author: Muzawir Arief, Dept of Health and Social Management, University of Eastern Finland, Kuopio Finland. E-mail: muzawia@uef.fi

Proper accessibility of online information related to health is a necessity for individuals who seek out health-related information due to their physical conditions.

Meanwhile, the World Wide Web Consortium (W3C) have designed the Web Content Accessibility Guidelines (WCAG) in collaboration with international figures and institutions to create international general standard of web accessibility that is suitable with any institutions, public and governments [7]. The guidelines have targeted web developers, web authoring tool developers, web accessibility evaluation tool developers, and others (individuals, decision makers, administrators, scholars. etc.) [7]. WCAG 2.0 has been approved as an ISO standard (ISO/IEC 40500:2012). WCAG2.0 encompasses web pages, web applications, mobile device applications, phones and tablets, digital TVs, wearables, devices in car dashboards and airplane seatbacks, household equipment, and other Internet of Things (IoT) [7], [8]. WCAG 2.0 have four accessibility principles: perceivable, operable, understandable, and robust, which must be fulfilled to avoid violations [9]. These principles have their own success criteria that determine the type and number of violations. The perceivable principle has 22 success criteria, the operable principle has 20, the understandable principle has 17, and the robust principle has 2. The UN prioritized web accessibility in resolution 61/106, and the standards have been implemented in their system [10]. With the resolution, access to information and communication technology have become basic human rights [11]. WCAG2.0 of W3C are the only internationally recognized standards [10]. Therefore some countries have ratified the standards following this UN recognition [12], [13].

This study aims to evaluate the effectiveness of the implementation of WCAG 2.0 in country with and without policy implementation. The study compared country that adopt the WCAG 2.0 standard in health information provision with the one that do not adopt the WCAG 2.0 standard to study the particularly of policy implementation on Alzheimer Disease (AD). To reach the aim, the study addressed two research questions (1) What is the level of web accessibility of AD-related information online both in the UK and Indonesia? (2) What are the similarities or differences in the information available online for AD between the two countries, one with a policy of accessibility (UK) and one without (Indonesia)?

2. Materials and Methods

Two countries were chosen for this case study. The United Kingdom was chosen for being a country that has adopted the WCAG2.0 standard [13], [14], and Indonesia was selected as a country that has not adopted the WCAG 2.0 standard. The selected countries were chosen also based on the first author language capacity. In addition, the idea to measure the difference have become consideration to choose the countries that are very contrasting based on their economy level besides the policy adoption. Web selection was based on the inclusion criteria of local language content: Bahasa Indonesia for Indonesia and English for the UK. It is important to note that the same word "Alzheimer" is used in both local languages to refer to AD [15]. The first ten websites from each country were picked based on their rank on the Google search engine list. The method for selecting the first ten websites has been previously used [16]. Google was first set up to the country-specific language, and a search was performed using the keyword "Alzheimer", due to its high frequent used among individuals searching for information related to AD [9], and it is known that AD affects cognition and memory [17], as consequences better accessibility of online information

is required. Among the obtained results, the website should be in a webpage format, and information from the same source will be excluded. The websites should have content related to AD, and the websites should not be from a journal publisher, university, or scientific research institution with the reason they are usually for academic community.

Website evaluations were performed for ten public online health information pages using the WCAG 2.0 evaluation tool recommended by W3C “Axe” [14]. In addition, Axe tool evaluate the website compliance to WCAG 2.0 with respect to WAI-ARIA (Web Accessibility Initiative-Accessible Rich Internet Application) 1.0 that is for assistive technology such as screen reader [18]. Based on the number of violations found, the website was assigned to one of four levels: low (0–20 violations), medium (21–50 violations), high (51–100 violations), and very high (above 100 violations). The term of violation means the unfulfilled WCAG2.0 criteria for accessibility for all. Statistical analysis was carried out using SPSS version 25. Descriptive and Mann-Whitney *U* tests were used to measure the differences between the two countries with and without WCAG 2.0.

3. Results

Entering the keyword “Alzheimer” for Indonesia and UK searches yielded 153,000,000 and 232,000,000 results, respectively. Using Axe tool analysis on the 20 websites (10 per country search), 17 types of violations were found. These were the violations found: (1) ARIA roles used must conform to valid values, (2) Elements must only use allowed ARIA attributes, (3) hidden ARIA element must not contain focusable elements, (4) Certain ARIA roles must contain particular children, (5) ARIA attributes must conform to valid values, (6) Buttons must have discernible text, (7) Element must have sufficient color contrast, (8) IDs used in ARIA and labels must unique, (9) ID attribute values must be unique, (10) Frames must have title attribute, (11) Html must have language attribute, (12) Image must have alternate text, (13) Form elements must have labels, (14) Links must have discernible text, (15) “ul” and “ol” must only directly contain li, script or template elements (16) Zooming and scaling must not be disabled, and (17) Timed refresh must not exist. Most website violations in the UK-specific search were under the low level category, whereas none of the violations in the Indonesia-specific search were of low level. Most of the tested websites in the Indonesia-specific search were found to have medium-level violations.

In addition, the 20 selected websites could be categorized into four types: non-government organization (NGO) websites (*n* = 5), medical related company websites (*n* = 4), newspaper-type websites (*n* = 10), and government agency websites (*n* = 1) (Figure 1 below).

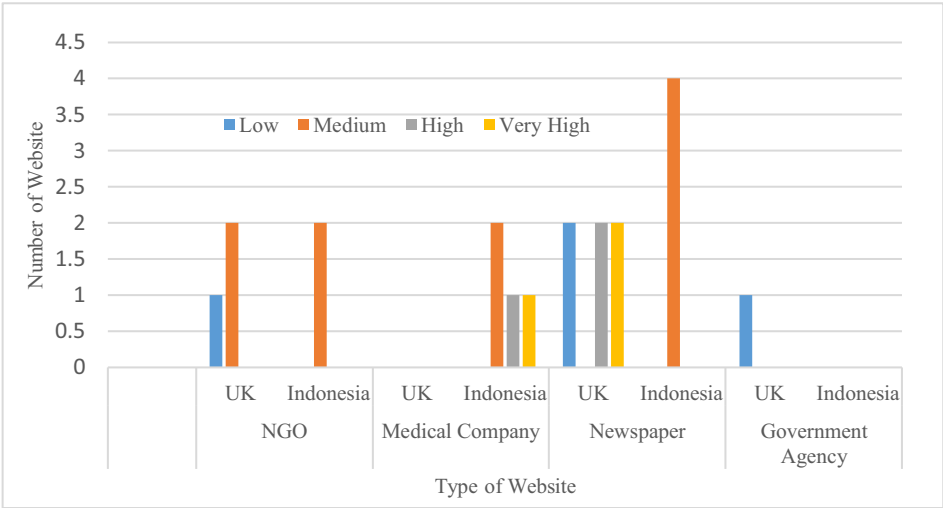


Figure 1. Violation Occurrences Based on Website Type

NGO and government websites in the UK showed good compliance. Newspaper-type websites were the most common type of website for both countries. In the UK, two of the six newspaper-type websites with AD information showed good compliance with the standard and two showed a very high level of violation. On the other hand, all four online news websites in Indonesia showed medium-level violations.

The Mann-Whitney test revealed no significant differences between countries with and without WCAG 2.0 standard implementation. In addition, the most dominant type of violations found for both countries was “element must have sufficient color contrast” and “links must have discernible text.” Furthermore, “ID attribute value must be unique” was the third most violated standard in UK webpages while “image must have alternate text” was the third most violated standard in Indonesian webpages. A similar result was found when comparing newspaper-type websites and other website types.

4. Discussion

Implementing web accessibility policy globally continues to be a challenge. Despite national regulations, for example the United States Section 508 regulations requires every government agency to comply with web accessibility, accessibility problems persist [19]. In addition, the WCAG 2.0 policy has no impact outside government agencies [20]. Even within the municipal government, there are many issues such as national support and close collaboration across local administrators and external parties, citizen and other rules [21]. For example, although in the UK better accessibility was on the health administration agencies, only NGOs have slightly implemented better accessibility. Meanwhile, online newspapers were found to be the worst online channel in terms of adopting the WCAG2.0 standard.

The complexity in deploying WCAG 2.0 is elaborate, e.g., lack of law enforcement, no report or certification of WCAG 2.0 compliance, limited resources, policy content, lack of responsibility, and unclear guidelines from the department that

released the regulation [21], [19], [20]. The study revealed that the most common mistakes that should be easy to be understood, such as alternate text and color contrast in webpages still exist means whether web developers have reluctance to comply or excessive working load with short time due to particular business characteristic.

Another factor that might influence the evaluation is the richness of website content. It is known that the richer of the website content the more potential violation might exist.

References

- [1] D. K. K. Wong & M. K. Cheung, Online health information seeking and ehealth literacy among patients attending a primary care clinic in Hong Kong: A cross-sectional survey, *J. Med. Internet Res* **21** (2019), e10831.
- [2] L. Silver, C. Huang & K. Taylor, In Emerging Economies, Smartphone and Social Media Users Have Broader Social Networks, Pew Research Center, 2019.
- [3] S. A. LaValley, M. T. Kiviniemi & E. A. Gage-Bouchard, Where people look for online health information, *Health Info. Libr. J* **34** (2017), 146–155.
- [4] S. Fox, E-patients With a Disability or Chronic Disease, Pew Research Center, 2007.
- [5] L. Nölke, M. Mensing, A. Krämer & C. Hornberg, Sociodemographic and health-(care-)related characteristics of online health information seekers: A cross-sectional German study, *BMC Public Health* **15** (2015), 31.
- [6] European Commission, Digital Inclusion and Skills, 2014.
- [7] W3C, Web Content Accessibility Guidelines (WCAG) Overview, 2018. Available from: <https://www.w3.org/WAI/standards-guidelines/wcag/>
- [8] W3C, Mobile Accessibility at W3C, 2019. Available from: <https://www.w3.org/WAI/standards-guidelines/mobile/>
- [9] T. K. Houston, J. E. Volkman, H. Feng, K. M. Nazi, S. L. Shimada & S. Fox, Veteran Internet Use and Engagement With Health Information Online, *Mil. Med* **178** (2013), 394–400.
- [10] United Nations, Accessibility Guidelines for United Nations Websites, Available from: <https://www.un.org/en/webaccessibility/index.shtml#GA>.
- [11] United Nations, Convention on the Rights of Persons with Disabilities, 2017.
- [12] US Government, Accessibility policy, 2019. Available from: <https://www.usa.gov/accessibility>.
- [13] UK Government, Equality Act 2010, National Archive, 2010. Available from: <http://www.legislation.gov.uk/ukpga/2010/15/contents>.
- [14] Education and Outreach Working Group, Web Accessibility Evaluation Tools List, 2018.
- [15] Cambridge Dictionary, Available from: <https://dictionary.cambridge.org/dictionary/english-indonesian/alzheimer-s-disease>.
- [16] L. Spagnoli, M. Navaro, P. Ferrara, V. Del Prete & F. Attena, Online information about risks and benefits of screening mammography in 10 European countries, *Medicine (Baltimore)* **97** (2018), 1–6.
- [17] M. Fernández, A. L. Gobartt & M. Balañá, Behavioural symptoms in patients with Alzheimer's disease and their association with cognitive impairment, *BMC Neurol* **10** (2010), 87.
- [18] Deque University, ARIA attributes must conform to valid names. Available from: <https://dequeuniversity.com/rules/axe/3.4/aria-valid-attr>.
- [19] A. Olalere & J. Lazar, Accessibility of U.S. federal government home pages: Section 508 compliance and site accessibility statements, *Gov. Inf. Q* **28** (2011), 303–309.
- [20] M. L. Leitner, C. Strauss & C. Stummer, Web accessibility implementation in private sector organizations: motivations and business impact, *Univers. Access Inf. Soc* **15** (2016), 249–260.
- [21] E. M. Velleman, I. Nahuis & T. van der Geest, Factors explaining adoption and implementation processes for web accessibility standards within eGovernment systems and organizations, *Univers. Access Inf. Soc* **16** (2017), 173–190.

Evolution of Interdisciplinarity in Medical Informatics in Europe: Patterns from Intertwining Histories

Casimir A KULIKOWSKI^{a,1}, George I MIHALAS^b

^a*Rutgers - The State University of New Jersey, New Brunswick, NJ, USA*

^b*Victor Babes University of Medicine and Pharmacy, Timisoara, Romania*

Abstract. The IMIA History project book we are co-editing with colleagues from the IMIA History Working Group includes histories of early contributions to medical and healthcare informatics, as described by a sample of pioneers and experts, detailing how their own ideas developed from their work on various topics in the field at the beginnings of their contributions to the field. Its contents serve as a preliminary guide for meta-analyses of how the different contributors state their personal interdisciplinary origins from today's perspectives. In this short article we provide a brief preview of how an analysis of disciplinary characteristics from individual histories can begin to shed light on processes of interdisciplinary evolution of medical informatics in Europe.

Keywords. Interdisciplinarity, History, Medical and Healthcare Informatics

1. Introduction

The history of science illustrates how the development of different disciplines and domains of study has almost invariably given rise to various “boundary sciences” or interdisciplinary fields in order to deal with problems through new approaches that draw from methods and threads of inquiry and experimentation from several existing disciplines. Such inter-weavings usually prove to be quite productive, yielding interesting new scientific discoveries or important technological and practical advances. A web of inter-relations between various domains frequently takes place, revealing sometimes-unforeseen links between apparently unrelated fields.

Usually, the very name of a new interdisciplinary domain tells us about its origin, specifying the “parent domains” whose intersection or blending has created the new discipline, as in biochemistry, psychopharmacology, and bioinformatics, among many others. This is also the case with medical informatics, which results from the intersection and interweaving of two very large fields of practice and study – medicine (relying on the biomedical and life sciences applied to healthcare) and informatics (relying on computer and mathematical sciences applied to information technology). Its name

¹ Corresponding Author, Rutgers - The State University of New Jersey, Piscataway, New Brunswick, NJ 08855, USA; E-mail: kulikows@cs.rutgers.edu

reflects its nature as an interdisciplinary field, or, as van Bommel stated: “medical informatics is interdisciplinary *avant-la lettre*” [1].

Studies on interdisciplinarity analyze various aspects of the links between concepts, knowledge and data structures or methods involved in the rise of a new domain. Such meta-analyses require a study and insight into the origins of the connections, which are not always presented clearly enough in the literature largely produced by authors from the existing disciplines. An incursion into a domain's history can cast some rays of light, revealing the roots of subsequent studies, which had strong follow-up and became substantive subfields of the new field.

2. Materials and Methods

We have recently carried out such a study, examining the roots of interdisciplinarity in medical informatics [3,4], based on 54 stories from European scientists who started their careers before 1980, or the early stage in medical informatics history, and which has been called its “pioneering period” [5]. The breakdown by length of experience in the field as measured from their year of university graduation consisted of 14 from 1960 and earlier, 20 from the 1961-1970 period, and 20 from 1970 to 1980. By gender, 48 were male and 6 were female.

We were interested in analyzing this data to uncover relations between the first professional background discipline of a pioneering investigator, through their professional affiliation, type of occupation or professional milieu within which the early medical informatics topics were approached (academic/university, healthcare organization, industry, other institution) and the medical informatics topics or approaches, both initial and subsequently developed.

3. Results

3.1. Professional Origins

Our study showed that 48 or about 81% of this sample of European pioneers in medical informatics had some kind of background involving technology or the exact sciences while 27 or 50% had backgrounds in medicine or the life sciences; with an overlap of 30% representing those who had a double specialization. Among the scientific or technical backgrounds, the most common were mathematics/statistics (19 persons), physics (13 total, but 10 of whom also had mathematics/statistics backgrounds also) and engineering (14 of which 7 involved computer science) but also chemistry/biochemistry (4) and biophysics (4). The biomedical profiles were predominantly in traditional medical specialties (20), while other fields included nursing (2), public health/epidemiology (2), dentistry (1) and psychology (1).

A obvious question that arises from this distribution of backgrounds is: “why is there such a high percentage of those coming from scientific and technical fields?” One might speculate that this might arise from an asymmetry that is observed in most interdisciplinary domains, including medical informatics, where one field provides the object of study (in this case biomedicine), while the others provide methods of analysis and interpretation (informatics). It has often been observed that those working in

information technology as researchers and scientists were able to foresee the great potential of computing technologies for various applications before the application-domain specialists come to fully realize and embrace it. The altruistic goals of medicine and nursing, and the desire to uncover the mysteries of life have led to a proliferation of several interdisciplinary fields, including medical informatics. Multidisciplinary teams have arisen, bringing together specialists from various fields trying to solve the complex problems that cross the boundaries between the more formal analytic disciplines, new technologies, and the field of medical and healthcare practices. But, to support effective collaboration and integration, specialization in a single domain is rarely enough. It is not a surprise that many researchers have felt the need for a double specialization. This has happened in both directions – medical and healthcare practitioners involving themselves in informatics and/or mathematics/statistics, while mathematicians, statisticians, engineers and computer scientists have in turn focused on biomedicine.

3.2. Occupational Backgrounds and Experiences

Most early medical informatics practitioners in our European sample (over 75%) were mainly academics who started their work within various departments of universities, with over 20% of them being from university hospitals; while others came from industry, non-academic hospitals or public health or research units.

It is interesting to note that we have found a large number (over 25%) involved in governmental activities or pan-European organizations. Indeed, the potential of information technology to be part of the process of improving healthcare activities often appear to have drawn the attention of official or governmental institutions (ministries or other healthcare authorities) from the earliest days of medical informatics.

The most frequent research activities involved coordinating or participating in various projects or acting as experts. Given the preponderance of academic affiliations, education and training was omnipresent, contributing significantly to the development of the field.

3.3. Subfields and Topics

One of the most interesting aspects of this first overview study drawing on histories of how experts developed their interests and contributions to the field has to do with the evolution of key problems addressed or topics defined: how the ideas evolved, what challenges presented themselves, and how they were overcome. Such inquiries are important since they better disclose the essence of those connections which motivate and build the effective interdependences and webs of interdisciplinarity. One can understand interdisciplinarity more fully as going beyond being just the passive application of methods from one field in a study to answer a direct question about some concept or object from another field. A simple statistical analysis of some clinical data can hardly be considered an interdisciplinary study. The substance of interdisciplinarity is a bilateral relationship between people from different disciplines who actively work together to ask new types of questions, with the goal of uncovering new features of a subject or object of study which can drive the development of new methods in order to reach a deeper understanding of the phenomenon being investigated. Such an approach was highlighted earlier by van Bommel in his detailed personal story [6], tracing the evolution of ideas in medical informatics as they changed over the course of his career.

In the present overview we use the four research challenge subfields or areas in medical informatics proposed by Kuhn et al. [6]: bioinformatics and systems biology, biomedical engineering and informatics, health informatics and individual healthcare, and public health informatics. As expected, since studies on bioinformatics and systems biology had barely started during the pre-1980 pioneering period we are examining, they represented only 5% in our sample, while studies on health informatics and individual healthcare are most plentiful, representing almost 45% of the total during this period. Studies in biomedical engineering and those on public health informatics had almost equal shares of approximately 25%, representing the disciplines that existed before informatics came to be defined as a separate field in its own right.

As discussed by Altman and colleagues recently [8], these subfields or areas of medical informatics interconnect at several levels and it is sometimes difficult to place studies in one specific group vs. another; so we have developed a number of criteria in order to do so, as discussed in more in detail in [3]. A better overall picture can be obtained by using a more granular, or refined breakdown of the areas or topics of study. The most frequently covered major topic out of a total of 65 was health information systems (28, or almost 20%), of which 14, or half, dealt with hospital information systems, with the others being clinical/laboratory or national information systems. The topic of data bases, most often associated with data processing or statistical analyses accounted for 15%, decision support systems including expert systems, AI and ontologies accounted for 10%, while electronic health records were the topic of another 10%. Remaining topics involved security and protection, standards and technological assessment, processing biological signals, medical imaging, modeling and simulation of pharmacological or biophysical phenomena, healthcare management and organizational impact, ethical issues and qualitative assessment studies, nursing informatics, clinical guidelines and telemedicine. Each were covered by a relatively small percentage of the persons involved in this sample from the early period of medical informatics research in Europe.

4. Discussion

Our study is preliminary, yielding just initial quantitative results, which, nevertheless point to how we can proceed with further more detailed follow-up studies. An important consideration is that personal narratives in free style do not lend themselves to the extraction of very precise data. Secondly, the professional backgrounds considered are very general and based on academic degrees or specializations, which does not characterize the important group of self-taught people, especially those who were self-instructed in the use of computers, which makes actual numbers of those with double specializations likely to be even higher than what is reported here. Since the boundaries between topics are not crisply defined; with some types of investigations distributed over several topics, the assignment of a study to a topic is not straightforward. Nevertheless, despite these limitations, the overall results are already quite suggestive about the patterns of backgrounds and interests of individuals, and their intertwinings, and correspond to the experiences of the authors in what they have observed over the years.

Some topics have persisted over time, remaining most relevant, or “hot”, either due to their generality, complexity and potential impact (artificial intelligence applications, the digital patient) or strongly determined by “external factors” (large healthcare information systems, requiring socio-economic and political involvement), or both (EHR

systems). Some topics have faded – such as biological signal analysis, where research has mostly migrated to specialized biomedical disciplines, or, once mature, to industrial/commercial development. Beyond the period covered by this analysis, totally new topics have arisen – having mostly to do with the internet, social media, privacy and security, patient empowerment, big data and, most recently, data science. These, while building on earlier methodologies and studies of the social impact of computers and information, have to confront completely new challenges of scale, complexity, and density of interactions in the evolving human-machine ecologies of today.

5. Conclusion

We can see from the histories or essays submitted to the IMIA history book that they give the reader selected glimpses of individual professional studies, investigations, and practices, which, taken together, allows us to begin to discern a map for tracing the ideas and practices of medical informatics history. A thorough analysis of the stories specific to various periods in medical informatics history, which will require careful definition and study, would yield a dynamic map representing “medical informatics in evolution” – a necessary enterprise which looks forward to more complete investigations and the involvement of new generations of biomedical and health informatics professionals.

Acknowledgements

The authors wish to thank the reviewers of the original submission for suggesting improvements. They also thank their friends in the IMIA History Working Group, the ACMI Committee of Historians, and other colleagues who have helped in the international biomedical and healthcare history projects related to this work.

References

- [1] J. H. van Bommel. Medical informatics is interdisciplinary avant la lettre. *Methods Inf Med* **47** (2008), 318-321
- [2] C. A. Kulikowski, G. I. Mihalas et al., 50th Anniversary International Medical Informatics Association (IMIA) History Working Group and Its Projects, *Stud Health Technol Inform* **245** (2017), 758-762.
- [3] G. I. Mihalas, C. Kulikowski. Roots of Interdisciplinarity in European Medical Informatics. *Stud Health Technol Inform.* **262**, (2019), 146-149
- [4] I. Masic, G.I. Mihalas (eds), Contributions to the History of Medical Informatics, Avicena, Sarajevo, 2014.
- [5] G. I. Mihalas, Evolution of Trends in European Medical Informatics. *Acta Informatica Medica* **22** (2014), 37-43.
- [6] J. H. van Bommel. People and Ideas in Medical Informatics. A Half Century Review. *Yearb Med Inform.* (2011), 175-82
- [7] K. A. Kuhn, A. Knoll, H. W. Mewes et al. Informatics and medicine - from molecules to populations. *Methods Inf Med* **47** (2018), 283-295.
- [8] R. B. Altman, R. Balling R, J. F. Brinkley et al. Commentaries on "Informatics and medicine: from molecules to populations", *Methods Inf Med* **47** (2018), 296-317.

General Data Protection Regulation (GDPR) in Healthcare: Hot Topics and Research Fronts

Farhad FATEHI^{a,b,1}, Farkhondeh HASSANDOUST^c, Ryan K. L. KO^a, and Saeed AKHLAGHPOUR^a

^a*The University of Queensland, Brisbane, Australia*

^b*Tehran University Medical Sciences, Iran*

^c*Auckland University of Technology, New Zealand*

Abstract. General Data Protection Regulation came into effect across the European Union in May 2018 but its implications in healthcare are yet to be fully understood. The aim of this study was to identify the fronts and hot topics in research on GDPR in healthcare. We analyzed the relevant records in Scopus through bibliometric and scientometric approach and visualization techniques. A set of 155 records was obtained and processed for co-occurrence analysis of key terms and concept mapping. The number of published papers showed a steep rise in the past two years, mainly by European countries. Analysis of the abstract of the papers showed that data protection, privacy, and big data were the most frequently used terms. Three dominant research fronts of GDPR are 1) general implications of GDPR, 2) technology aspects of GDPR, and 3) GDPR in healthcare service. Blockchain and machine learning are among the reemerging topics of GDPR research.

Keywords. GDPR, General Data Protection Regulation, bibliometrics, scientometrics

1. Introduction

With the proliferation of invasive digital technologies and the emergence of data-exploiting business practices, it is increasingly difficult for individuals to maintain control over their own personal data. Consequently, this issue of control over personal data has become a significant subject in European privacy law. Compared to earlier regulations, the General Data Protection Regulation (GDPR) explicitly addresses the rights of individuals concerning this issue [1]. This means that organizations are no longer able to use individuals' personal data without their consent. GDPR dictates that entities collecting and processing data related to European Union (EU) residents adhere to GDPR articles regardless of where these entities are located, or where the data is stored. GDPR compliance support the protection of privacy at various levels. Healthcare data breaches not only can affect substantial adverse personal and social impacts on patients, but also suffer a huge cost [2]. Most notably for the healthcare community, GDPR

¹ Corresponding Author: Farhad Fatehi, The University of Queensland, Brisbane, Australia; E-mail: f.fatehi@uq.edu.au.

explicitly introduces regulations for “data concerning health”, “genetic data”, and “biometric data” [3].

In spite of the fact that GDPR is a recently developed legal instrument, there have been many research studies on GDPR since its adoption in 2016, such as [4; 5]. Thus, it is important to identify and monitor the research fronts and trends relating to GDPR for better understanding and integration of this regulation with other security, privacy, and technological policies. Research fronts refer to the most dynamic research areas in science and technology that attract the most scientific interest [6] and are the body of articles that researchers actively cite in a short time span [7]. One main characteristic of emerging research fronts is high citation in a short time interval. The objective of distinguishing hot topics within a research front is to identify significant insights that are likely to determine the future direction of research. This study aimed to establish a GDPR research agenda by identifying the current trends, emerging research fronts, and hot topics in research on GDPR in healthcare using a bibliometric analysis approach.

2. Methods

We conducted a bibliometric and scientometric analysis of the papers that have discussed GDPR in healthcare. The data were obtained through an electronic search of Scopus database in October 2019. No limit on year of publication or language was imposed. The results of the electronic search were exported and loaded into Bibliometrix package for R v.3.6.1 and VOSviewer v.1.6.13 for bibliometric and scientometric analyses and visualization. [8; 9]. Research fronts and hot topics were identified by concept mapping of the abstracts as well as co-occurrence analysis of the keywords of each record. We used co-occurrence analysis of keywords to extract the topics of research and find the links between these topics directly from the content of abstracts. Sankey diagram was used to visualize the main items of three fields (i.e. Country, Keywords, Year), and how they are related.

3. Results

The first paper on GDPR in healthcare has published in 2012 and the number of publications was less than 10 per year until 2017, when it shows a radical increase up to 76 papers in 2018 (the year that GDPR was enacted). Germany has published the highest number of papers, followed by the UK and the Netherlands (Figure 1).

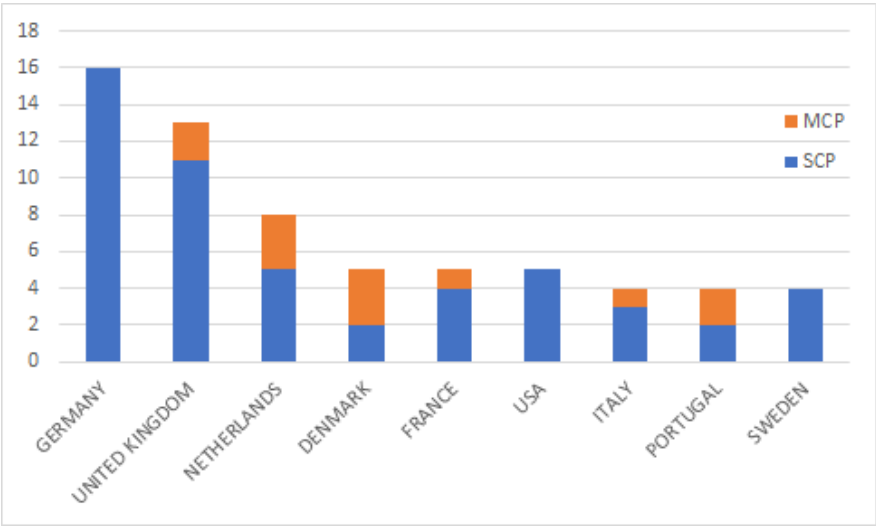


Figure 1. Countries with more than two papers based on the corresponding author’s address. (MCP: Multiple-country publication; SCP: Single-country publication)

A three-field Plot (Sankey diagram) of Country, Keyword, and Year of publication of the cited references was created to depict the proportion of research topics for each country and the recency of the papers that they cited. As shown in Figure 2, the main interests of GDPR researchers in France are big data and data protection. The dominant topic of research in the Netherlands is privacy. Most of the papers that discussed consent have published by Ireland, the Netherlands, and the UK. Research on GDPR and Blockchain, though are few in numbers, are mainly published by USA, Norway, UK, and Netherlands.

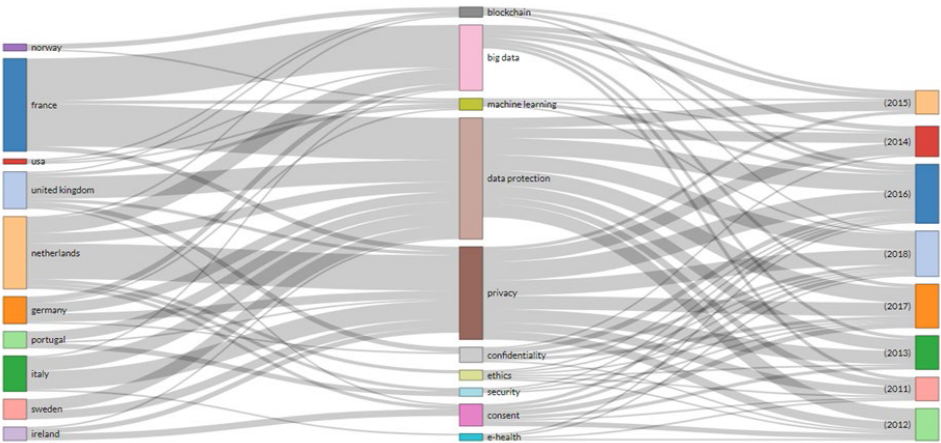


Figure 2. A three-field Plot (Sankey diagram) of Country, Keyword, and Year of publication of the cited references for the ten most researched topics.

The co-occurrence map of the most frequently used terms in the abstract of the papers was created to reveal the link between the terms and visualize the main clusters

Table 1. Research fronts and hot topics in GDPR research

Cluster (Color in Figure 3)	Research Front	Hot Topics
Cluster 1: Legal (Red)	General Implications of GDPR	Personal Data, Legal Aspects, Data Protection, Data Subject, Data Subject Rights, Data Access, Genetic Data, and Context-Specific Applications
Cluster 2: Technical (Green)	Technology Aspects of GDPR	Information Systems, Technology Solutions, Risks, Security, Controls, Timeliness of Data
Cluster 3: Health (Blue)	GDPR in Healthcare Practice	Patient, Patient Consent, Hospital Implementation, Health Data, Data Quality, and Use of Data

Prior evidence tells us that these dynamic research areas that create the greatest interest among the scientific community tend to be the most promising in terms of generating highly cited research, creating attitudinal change among researchers, challenging existing orthodoxies, anticipating possible paradigm shifts, attracting institutional funding, and enhancing the business value of the outcomes [4].

Researchers interested in data protection in healthcare can use our findings to plan for studies on the identified hot topics by observing the research trends, as well as research gaps in this domain. Likewise, journal editorial boards can benchmark their respective journals against these findings and introduce policies for publishing research on the increasingly important topic of personal health data protection. We believe Figure 2, in particular, reveals several opportunities for proposing special issues on important yet understudied topics, such as implications of GDPR for machine learning, Blockchain applications in healthcare, and ethics of collecting and processing personal health data.

References

[1] Van Ooijen, I and Vrabec, HU, Does the GDPR enhance consumers’ control over personal data? An analysis from a behavioural perspective, *Journal of Consumer Policy* **42** (2019), 91-107.

[2] Iyengar, A, Kundu, A, and Pallis, G, Healthcare informatics and privacy, *IEEE Internet Computing* **22** (2018), 29-31.

[3] McCall, B, What does the GDPR mean for the medical community?, *The Lancet* **391** (2018), 1249-1250.

[4] Albrecht, JP, How the GDPR will change the world, *Eur. Data Prot. L. Rev.* **2** (2016), 287.

[5] Wachter, S, Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR, *Computer Law & Security Review* **34** (2018), 436-449.

[6] Upham, S and Small, H, Emerging research fronts in science and technology: patterns of new knowledge development, *Scientometrics* **83** (2009), 15-38.

[7] Shibata, N, Kajikawa, Y, Takeda, Y, and Matsushima, K, Detecting emerging research fronts based on topological measures in citation networks of scientific publications, *Technovation* **28** (2008), 758-775.

[8] Van Eck, N and Waltman, L, Software survey: VOSviewer, a computer program for bibliometric mapping, *Scientometrics* **84** (2009), 523-538.

[9] Aria, M and Cuccurullo, C, bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics* **11** (2017), 959-975.

[10] Tikkinen-Piri, C, Rohunen, A, and Markkula, J, EU General Data Protection Regulation: Changes and implications for personal data collecting companies, *Computer Law & Security Review* **34** (2018), 134-153.

Health ‘Big Data’ Value, Benefit, and Control: The Patient eHealth Equity Gap

Paul DEMURO^{a,1}, Carolyn PETERSEN^b, Paul TURNER^c

^a*Nelson Mullins Broad and Cassel, Fort Lauderdale, FL, USA*

^b*Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA*

^c*eHealth Services Research Group (eHSRG), ICT, University of Tasmania, Australia*

Abstract. Improvements in artificial intelligence and machine learning combined with the availability and exponential growth in individual and population health data offer opportunities for innovative patient/citizen-centered eHealth solutions. However, this confluence of social, technical, and economic interests can result in the privatization of control of patient data and contribute to widening inequity in access to healthcare. This paper explores these issues and advocates for a more equitable approach to advances in health big data for patients and citizens.

Keywords. Health disparities, big data, eHealth, health equity, digital divide, artificial intelligence, patient-centered

1. Introduction

Recent developments in artificial intelligence (AI) and machine learning (ML) have combined with the availability of and exponential growth in individual and population health data to offer opportunities for new and innovative patient-centered eHealth applications [1] eHealth has been defined as “the cost-effective and secure use of information technologies (ICT) in support of health and health related fields [2]” mHealth may be seen as the use of mobile devices for medical and public health purposes [3]. This convergence has many positives, including strengthening the patients’ voice by enabling them to share data both with one another and with health professionals, which enables being more active participants in their own care [4].

Simultaneously, the health information technology (HIT) industry views health big data and patient-generated health data (PGHD) as an opportunity to generate value through the commodification of personal and population to generate value through the commodification of personal and population data aggregation, mining and exploitation. Opportunities for HIT to benefit from smart sensors, applications, and services abound as AI automates “datafication of health” [5], problematizes individual de-identification [6], and contributes to erosion of trust in health data sharing [7].

The capacity of the private sector to exert ever greater control over this big data aided by advances in AI and ML is little understood by individuals and policy-makers. It is evident that data commodification is occurring in ways that raise social, ethical,

¹ Corresponding Author, Paul R. DeMuro, Nelson Mullins Broad and Cassel, Fort Lauderdale, FL USA; e-mail: paul.demuro@nelsonmullins.com

and moral questions [8] and debate about value and benefit [9]. Failure to appreciate the implications of these technologies on health data hampers discussions about data privacy, data ownership, and consent for access and use of personal data. Given the pace of technological change, it is unsurprising that legal and policy frameworks, as well as mechanisms for their enforcement, remain inadequate. Technology-agnostic responses that meaningfully address data ownership, data protection, data retention and transmission, and data control processes are limited in scale and scope [10].

These socio-technical and economic developments may widen existing disparities in healthcare access between empowered and disempowered individuals and groups in healthcare access between empowered and disempowered individuals and groups [11]. Regardless of benefits that may accrue to some people, others are further marginalized. Many people lack the technical, health and genetic literacy to benefit from opportunities driven by big data and related patient-centered care models [12]. This paper explores these issues and advocates for a more equitable approach.

2. Background

Public and private health sector agencies are increasingly reliant on digital platforms for service delivery. These platforms are themselves become increasingly complex as AI, ML, and the Internet of Things (IoT) become embedded in HIT [13]. In combination, these developments mask the very real challenges that many individuals and groups face in accessing, using and benefiting from eHealth [14-16].

The General Data Protection Regulation (GDPR) provides for the protection of natural persons with respect to the processing of personal data and its free movement [17]. It applies to personal data relating to an identifiable person who can be identified in particular by reference to an identifier [18], such as those seeking health services.

Many patients find it challenging to understand how data held on them within eHealth systems is being managed, controlled, and valued. While many citizens believe that they own their data, numerous health entities – providers, institutions, HIT vendors, and third parties – may also claim ownership [19]. Although many patients may not own their electronic health record (EHR) [3], they should have to consent to its use and have access to it, as the GDPR provides for same [17]. Even when patients perceive that ownership of a digital device may support their claim to data ownership, others still may control and monetize that data [20], such as via control of PGHD that is accessible by device manufacturers, app developers, and third-party trackers [21].

In this context the eHealth inverse care law (the potential for meaningful use of appropriate eHealth is inversely proportional to the needs/capacities of the individual or population served) and equity gap become most evident [22]. Socially disadvantaged people are already known to be disproportionately affected by consent issues, and this is even more problematic in eHealth domain [23]. In the EU, however, such consent must be specific and informed, should cover all processing purposes [18] and requires a positive opt in [17]. Special category or certain sensitive data may be afforded more protection, e.g., in the healthcare context, genetics, biometrics, health, sex life, or sexual orientation [3], [17], [18]. Given that the GDPR is seen as more comprehensive than other such laws, including those addressing such issues in the United States, the GDPR provides for data protection in the context of international transfers where that is not deemed to be appropriate safeguards in those countries [17], [18]. Many digital

apps require high levels of technical and health literacy [24], which widens disparities within the eHealth ecosystems [25].

3. Data Sharing and Challenges for Patients and Citizens

There are huge opportunities and benefits associated with deployment of AI and ML into eHealth. However, the health care system needs a better understanding of how new technologies are best deployed. Equity, not equality, should guide roll-out and use.

Patient-focused eHealth technologies typically aim to obtain consent for data sharing in a “one-size fits all” approach that provides the same options to all users of the system. However, users have differing needs, experience, and skills, and few may experience equivalent risks when agreeing to secondary use of their data. Socially disadvantaged people are likely to be disadvantaged as users and sharers of personal health information, including PGHD.

The asymmetry of costs and benefits regarding data sharing has multiple aspects and, potentially, numerous adverse effects on health outcomes and quality of life. People with lower socioeconomic status may be less able to reap the benefits of treatments and products developed using PGHD shared by themselves and others. They may be unable to pay for care or medications, or may lack transportation to clinical sites where advanced treatments are available, or have greater difficulty understanding consent forms and digital products’ “Terms of Service,” and thus be at a disadvantage in making informed decisions and in advocating for themselves. Such asymmetry also may result in skewed research results affecting populations. For example, those with more education report greater willingness to sign up for biobank research [26-27], that may result in development of inaccurate clinical guidelines and ineffective policy-making that further extend eHealth inequities. The GDPR can provide some protections for covered individuals.

People who have uncommon traits may be identifiable if certain other pieces of information (e.g., birthdate, postal code) are known or can be located using public databases [28]. Even without the intent to share data without consent, the recurrence of data breaches leaves individuals exposed [29].

At the same time, individuals who have a high expectation of benefiting from research in which they participate will provide broad consent for use of their data in research without knowing exactly what that research will address [30]. Some expect min-health-specific benefits, such as improved patient-clinician communication [31], or view data sharing as a way to access better more personal information and care [32]. However, expectations of benefits cannot be guaranteed, and may contribute to the growing mistrust of healthcare organisations, clinicians, and the eHealth system. Although the GDPR provides some protections for secondary data use and the need for specific and informed consent, those who seek to benefit from such research may well interpret such protections broadly, such as contending that further processing of such data for specific scientific purposes is compatible with lawful processing operations.

4. Improving Use of Health Big Data for All Patients

Equity is about fairness. AI and ML have the potential to deliver precision medicine where “n equals you.” It is thus reasonable to require equity as a design principle for

new eHealth products and services so that patients and citizens can easily understand the benefits, value, and control of data manipulated by new tools and consented to their use when alone or in interaction with clinicians [33]. To reduce barriers, developers of mHealth apps, creators of health IT, policy-makers and researchers should consider citizens' psycho-social and economic contexts when using eHealth and avoid design that could compound disadvantage, bias, and inability to access resources.

5. Conclusion

This paper has highlighted how advances in AI and ML in eHealth may contribute to the privatization of control over patient data in ways contributing to widening inequity in access to healthcare. The paper has advocated for a more equitable and informed approach to emerging advances in health big data for all citizens.

References

- [1] E.J. Topol, High-performance medicine: the convergence of human and artificial intelligence, *Nature Medicine* **25** (2019), 44–56.
- [2] World Health Assembly. Resolution WHA58.28: eHealth. Ninth plenary meeting, 25 May 2005 – Committee A, seventh report. <https://www.who.int/healthacademy/media/WHA58-28-en.pdf?ua=1>.
- [3] World Health Organization. Global diffusion of eHealth: making universal health coverage achievable. Report of the third global survey on eHealth. <https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf;jsessionid=A7C841588C3B687B4745DDF419C4E01E?sequence=1>.
- [4] C. Petersen, Patient informaticians: turning patient voice into patient action, *JAMIA Open* **1** (2018), 130–135.
- [5] M. Ruckenstein, N. Dow-Schüll, The datafication of health, *Annual Review of Anthropology* **46**, (2017), 261–278.
- [6] H. Nissenbaum, H. Patterson, Biosensing in context: health privacy in a connected world. In *Quantified: Biosensing Technologies in Everyday Life*, The MIT Press, Cambridge, Massachusetts, 2016.
- [7] M.P. Tully, L. Hassan, M. Oswald, J. Ainsworth. Commercial use of health data—a public “trial” by citizens' jury, *Learning Health System* **3** (2019), e10200.
- [8] S. Sterckx, V. Rakic, et al, ‘You hoped we would sleep walk into accepting the collection of our data’: controversies surrounding the UK care.data scheme and their wider relevance for biomedical research, *Medicine, Health Care and Philosophy* **19** (2016), 177–190.
- [9] Z. Sanyal, P. Stolee, D. Juzwishin, D. Huserau, Economic evaluations of eHealth technologies: a systematic review, *PLoS ONE* **13** (2018), e0198112.
- [10] P. Balthazar, P. Harri, A. Prater, N.M. Safdar, Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics, *Journal of the American College of Radiology* **15** (2018), 580–586.
- [11] R. Pierce, Technology-driven “disparities” and technological solutions, *The American Journal of Bioethics* **16** (2016), 48–50.
- [12] C. Showell, P. Turner, The PLU problem: are we designing personal ehealth for people like us? *Studies in Health Technology and Informatics* **183** (2013), 276–280.
- [13] R. Singh, Ownership of healthcare data in the IoT era, part 1, *Compliance Today* (2019), 67.
- [14] D. Mangin, et al, Multimorbidity, eHealth and implications for equity: a cross-sectional survey of patient perspectives on eHealth, *BMJ Open* **9** (2019), e023731.
- [15] G. Iacobucci, Digital health: GPs aren't “Luddites” but want safe, equitable care, *BMJ* **364** (2019), 1258.
- [16] J. Adler-Milstein, et al, Electronic health record adoption in US hospitals: the emergence of a digital “advanced use” divide, *Journal of the American Medical Informatics Association* **24** (2017), 1142–1148.

- [17] European Union. Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016. Available at <https://op.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1/language-en>.
- [18] Information Commission's Office. Guide to the General Data Protection Regulation (GDPR). Available at <https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf>.
- [19] R. Sharma, Who really owns your health data? (2018), Available at <https://www.forbes.com/sites/forbestechcouncil/2018/04/23/who-really-owns-your-health-data/#2131e8f06d62>.
- [20] P.R. DeMuro, C. Petersen, Managing privacy and data sharing through the use of health care information fiduciaries, *Studies in Health Technology and Informatics* **265** (2019), 157-162.
- [21] R. Singh, Ownership of healthcare data in the IoT era, part 2, *Compliance Today* (2019), 65.
- [22] M. McCartney, Health technology and the modern inverse care law, *BMJ* **362** (2018), k3118.
- [23] N.P. Terry, Regulatory disruption and arbitrage in health-care data protection, *Yale Journal of Health Policy, Law, and Ethics* **17** (2017), 147.
- [24] B. Smith, J.W. Magnani, New technologies, new disparities: the intersection of electronic health and digital health literacy, *International Journal of Cardiology* **292** (2019), 280-282.
- [25] L. Kayser, A. Kushniruk, R.H. Osborne, O. Norgaard, P. Turner, Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: a framework for understanding users' needs, *JMIR Human Factors* **2** (2015), e9.
- [26] S.C. Sanderson, et al, Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US, *American Journal of Human Genetics* **100** (2017), 414-427.
- [27] C. R. Rogers, M. J. Rovito, M. Hussein, et al. Attitudes toward genomic testing and prostate cancer research among Black men. *American Journal of Preventive Medicine* **55** (2018), S103-111.
- [28] L. Sweeney, Only you, your doctor, and many others may know, *Technology Science* (2015), Available at <https://techscience.org/a/2015092093/>.
- [29] H.T. Tavani, et al, Responding to some challenges posed by the re-identification of anonymized personal data, *Computer Ethics - Philosophical Enquiry (CEPE) Proceedings* (2019), Article 2
- [30] I. Petersen, et al, Informed consent, biobank research, and locality: perceptions of breast cancer patients in three European countries, *Journal Empirical Research on Human Research Ethics* **9** (2014), 48-55.
- [31] M. Bradley, et al, Patients' motivations and interest in research: characteristics of volunteers for patient-led projects on PatientsLikeMe, *Research Involvement and Engagement* **2** (2016), 33.
- [32] O.A. Zarate, J.G. Brody, P. Brown, et al, Balancing benefits and risks of immortal data: participants' views of open consent in the Personal Genome Project. *The Hastings Center Report* **46** (2016), 36-45.
- [33] S. Aminololama-Shakeri, J.E. López, The doctor-patient relationship with artificial intelligence, *American Journal of Roentgenology* **212** (2019), 308-310

Human-Induced Errors in Networked Healthcare Research: Risk Management Under the GDPR

Jonas BIENZEISLER^{a,1,2}, Hauke FISCHER^{b,2}, Volker S. THIEMANN^b,
and Rainer RÖHRIG^a

^a*Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Germany*

^b*Division for Medical Informatics, University of Oldenburg, Germany*

Abstract. Modern research projects in healthcare research and medical research are oftentimes multi-centered, multi-disciplinary, and conducted by a consortium of multiple collaborators. Increasingly, the resulting data emanating emerging from different primary and secondary sources is linked on a personal level. The General Data Protection Regulation regulates many fundamental processes in such research projects. Despite the regulatory framework given, it can happen that a collaborator does not handle data properly. We propose a systematic risk management for the handling of data as well as a systematic error management.

Keywords. privacy, risk management, organization and administration, legislation and jurisprudence

1. Introduction

Benefiting from healthcare research, modern health information technologies can reduce human errors, improve safety, and thus refine the treatment provided to patients [1]. Aside from occurring during the delivery of care, human errors also ensue within healthcare research [2], inter alia when sensible data is processed in study projects. A common type of errors are *data breaches* when proprietary data is disclosed to an unauthorized viewer [3]. While this is not an immediate threat to most patients, the implications can be severe as well; sensible information mandates adequate handling [4].

Networked Healthcare research projects are conducted by a consortium of multiple collaborators. The data results from various primary and secondary sources and is linked on a personal level which opposes complete anonymization [5]. The handling of data relatable to a natural person is regulated by the *General Data Protection Regulation* (GDPR) [6]. The GDPR requires an effective risk management in the form of a *Data Protection Impact Assessment* (Article 34, GDPR [6]) and appropriate *technical and organizational measures* (Article 25 (1), GDPR [6]).

¹ Corresponding Author, Jonas Bienzeisler, Institute of Medical Informatics, Faculty of Medicine, RWTH-Aachen, Pauwelsstr. 30, 52074 Aachen Germany, E-Mail: email: jbienzeisler@ukaachen.de

² Shared first Authorship

However, when working as a trusted third party in networked research projects, we frequently encounter situations in which proprietary data are not handled adequately. One collaborator sent Patient IDs to the wrong collaborator, another revealed a password in email conversations, and hospitals shared medical information together with identifying data. Many researchers seem to be insecure when handling data breaches that are caused by human errors. Contrary to breaches arising from cyber-attacks, theft, loss or data leaks, these are understood to be ultimately inescapable [1,7]. In the case that an error leads to a data breach, a systematic risk- and error management with appropriate counteractions is necessary.

Despite the regulatory framework given by the GDPR, there is no easy-to-use risk and error management framework to systematically aid researchers in the event of a data-breach. We provide an abstract case example, from which we start to discuss methods and strategies that can be used to avoid, and if necessary, address risks or errors occurring in healthcare research projects.

2. Method

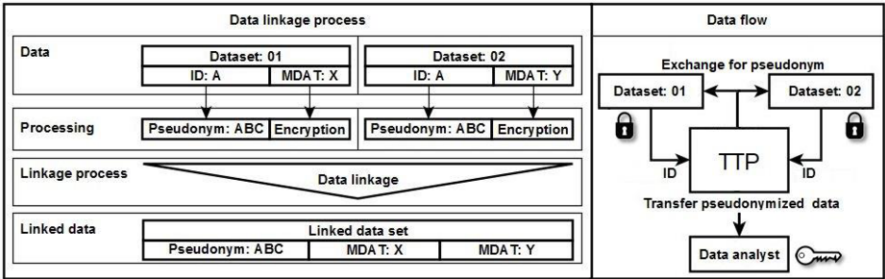


Figure 1: Data linkage process and data flow for the exemplary study design

We suggest tools to handle and prevent data breaches in health research projects when applying the current legislation regarding the GDPR in Germany [8]. Based on a generic study design, we discuss human errors and the compulsory risk management.

The exemplary research design is used to send two datasets collected by different collaborators via a trusted third party (TTP) to a third collaborator that evaluates the data. The TTP employs some method of data linkage to allow for the analysis of two medical datasets from two different sources (c.f. Fig.1). The datasets will be linked on patient-level through the pseudonymization of a project-wide, unique ID. This unique identifier could, for instance, be a case number identifying individual patients across datasets. The data is locally preprocessed with a software that encrypts only the medical data, not the ID. All processed datasets are then send to the TTP, which exchanges the ID for a new random pseudonym using a software that locally stores related IDs and pseudonyms. All data is then delivered to a data analyst as a file that contains data in a table-based structure. After decryption, each row contains medical data under the given pseudonym for analysis.

Human-induced data breaches in this context can happen during processing, data linkage, transmission or evaluation. One can discriminate three types of errors leading to a data breach event. Data can be (1) directly compromised (e. g. a screenshot of sensitive data in an email or data sent to an unauthorized person), (2) indirectly compromised (e. g.

not sufficiently pseudonymized) or (3) compromised due to misuse of technology (e. g. the wrong column in a dataset is pseudonymized).

For the assessment and handling of data breaches in such networked research, there are various laws to be considered. Aside from the GDPR, the BDSG, defines specific regulation for Germany. Processing of medical data for the purpose of research in both regulatory frameworks (Article 5, Article 9, GDPR [6]) and the BDSG (§22, §27, BDSG [8]). The GDPR requires a data protection impact assessment for data that could cause high risk to freedoms or rights of a natural person, and offers a minimum structure for such (Article 35 (1), (7), GDPR [6]). It is mandatory to employ *data protection by design* and *by default* similar to the *Privacy by Design* concept suggested by Cavoukian et al. [9]. The data controller (Article 4 (7), GDPR [6]) has to implement appropriate technical and organizational measures (Article 25 (1), GDPR [6]). Thus, the compromization of critical data requires the data controller to notify the responsible supervisory authority and affected persons within a determined period of time (Article 33, Article 34, Recital 85-88, GDPR [6]; §29, BDSG [8]).

Considering the GDPR, human errors have to be addressed on three levels; it is necessary to (1) work with adequate technical and organizational measures, (2) employ a systematic risk management, and (3) use an effective error management strategy

3. Results

Technical and organizational measures should render errors as unlikely as possible. State-of-the-art encryption and transport-encryption of all data are an indispensable prerequisite. In the context of the given example study, data need to be integrated button-up in the organizational design. Dataflow and software are usually designed with requirements engineering, so that it is –in theory– not possible to directly compromise data if software is used properly. Indirectly compromised data and the misuse of software have to also be addressed; the software in our example could check if the column containing the ID is not encrypted but the columns containing the medical data.

Direct and indirect compromization need to be conceptually addressed on an organizational level. Checklists and standard operating procedures (SOP) are generic methods to support processes. Checklists specify a set of goals to be fulfilled giving the user a framework to handle a specific process. The design of a checklist is challenging; a checklist must take the needs of different stakeholders into consideration and requires a careful mixture of guidance and independence, as not to limit people in their opportunity to make reasonable decisions [10]. SOPs give a detailed list of procedural steps which must be fulfilled. SOPs can be used for fixed processes that are repeated multiple times. In the context of our exemplary study, organizational measures would enhance and secure the project on working process level. SOPs would guide through the pseudonymization process, even when software is used. Another SOP would forbid workers revealing using the ID in email conversations. Checklists would offer the possibility to double check manually, whether the files have been processed according to the given framework conditions.

For a *systematic risk management* we adopted the HIT risk management concept from the IMIA Working Group for Health Informatics for Patient Safety [11]. During *Risk Assessment*, the risk of data breaches is analyzed and evaluated, *Risk Monitoring* serves as the instrument to identify data breaches and *Risk Control* comprises all counter

measures to manage data breaches on an operational level. Hence, risk management for data breaches is an iterative process resulting in continuous enhancements (cf. Fig. 2).

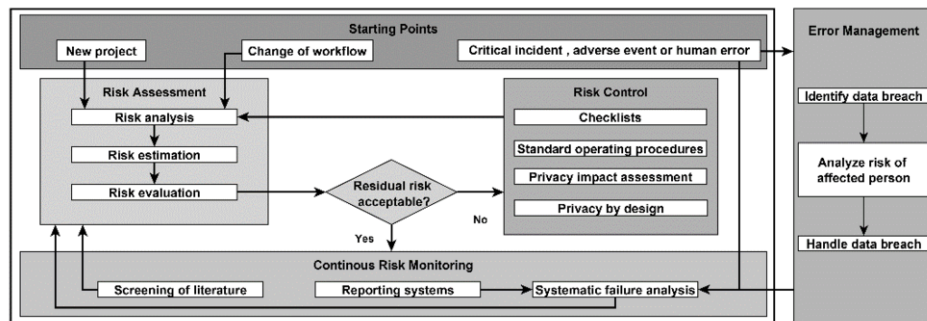


Figure 2: Systematic risk and error management of human-induced data breaches

Each error leading to a data breach has to be analyzed and handled individually as part of an *error management*. On an operational level, we suggest to divide error management into three generic phases and a subset of generic subtasks. The process should start immediately after the breach was discovered (Recital 87, GDPR [6]). In a first step the circumstances of current data breach are identified. It is crucial to categorize the compromised data in order to evaluate the risk. The GDPR defines different types of personal data into directly identifiable, pseudonymized, and anonymized [12]. In the second phase, the situation has to be evaluated in cooperation with the local data protection officer. The assessment of a risk for an affected person is the central process in handling data breaches. The target is to evaluate, if persons affected by the data breach will suffer a high risk in rights and freedom (Article 33 (1), GDPR [6]). If so, it induces a notification of the local authorities and the affected person. The notification should be made immediately and preferably within 72 hours (Article 33 (1), GDPR [6]). In the third phase, the actual data breach is handled by notifying authorities – if necessary – and isolating and deleting the compromised data. For the explicit procedure, the local data protection directives need to be considered.

What happens if the data analyst from our example received the full name of one participant and the respective ID when a data subject access request was handled? After identifying the data breach, she has to consult the local data protection officer to estimate the risk for the affected person. It is not mandatory to notify, when technical and organizational measures deny unauthorized persons access to the compromised data (Article 34 (3) (a), GDPR [6]). In our example this is the case, since the encryption denies the access to the medical data to any person except the data analyst; few people have access to the data and only the TTP can relate an ID to a natural person. Further, it is not obligatory to notify when the data leak will not result in a high risk to the rights and freedoms of data subjects (Article 34 (1), Article 34 (3) (a), GDPR [6]). If the data analyst analyzed the data breach, identified and isolated the compromised data, she can argue that the subsequent risk is low if the request containing name and ID is deleted (Article 34 (3) (a), GDPR [6]). We recommend documenting the data breach precisely; a controller has to be able to demonstrate a minimum risk (Recital 85, GDPR [6]). The consequential systematical failure analysis will then lead to a revision of potential SOPs for the handling of data subject access requests.

4. Discussion and Conclusion

Human-induced errors will always happen. We discovered the need for the development of adequate risk and error management in our everyday work. The framework of the GDPR gives room to interpretation and the legal space has yet to be embellished, which opposes easy handling of risks and errors. We developed a risk management concept and a protocol for error management giving guidance dealing with human induced errors and data breaches. Such a framework can prevent human-induced errors due to continuous learning and refined processes within distributed health care research projects.

The premise of all measures taken to reduce the consequences of human error is maximum transparency. A culture which facilitates the reporting of data breaches, which is the ultimate prerequisite for effective risk and error management, has been consensually established in medicine within the past decade [13]. However, it has not been translated to healthcare research yet. Reliability, safety, quality, and human factors are well-established fields in applied medicine and healthcare but not been applied on the level of healthcare research. A reporting system like the Critical Incident Reporting System (CIRS) [14] could be used to anonymously collect data of incidents and almost occurred incidents in order to establish an error culture. This data would contribute to developing and enhancing tools for a systematic risk- and error management.

5. REFERENCES

- [1] L.T. Kohn et al., eds., *To Err is Human: Building a Safer Health System*, Washington (DC), 2000.
- [2] J. Reason, Understanding adverse events: human factors. *Quality and Safety in Health Care* **4** (1995), 80–89.
- [3] M. Chernyshev et al., Healthcare Data Breaches: Implications for Digital Forensic Readiness. *J Med Syst* **43** (2018), 7.
- [4] W.N. Price and I.G. Cohen, Privacy in the age of medical big data. *Nat Med* **25** (2019), 37–43.
- [5] S. March et al., Quo vadis Datenlinkage in Deutschland? Eine erste Bestandsaufnahme. *Gesundheitswesen* **80** (2018), e20–e31.
- [6] European Union, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation): GDPR*, May 2018.
- [7] L.L. Leape and D.M. Berwick, Five years after To Err Is Human: what have we learned? *JAMA* **293** (2005), 2384–2390.
- [8] G. Recht, ed., *Bundesdatenschutzgesetz (BDSG)*, G. Recht, Merseburg, 2014.
- [9] A. Cavoukian et al., Remote home health care technologies: how to ensure privacy? Build it in: Privacy by Design. *IDIS* **3** (2010), 363–378.
- [10] Ø. Thomassen, et al., Implementation of checklists in health care; learning from high-reliability organisations. *Scand J Trauma Resusc Emerg Med* **19** (2011), 53.
- [11] E. Borycki, et al., Methods for Addressing Technology-Induced Errors: The Current State. *Yearb Med Inform* (2016), 30–40.
- [12] A.P. Olimid et al., Ethical approach to the genetic, biometric and health data protection and processing in the new EU General Data Protection Regulation (2018). *Rom J Morphol Embryol* **59** (2018), 631–636.
- [13] A.F. Smith and E. Plunkett, People, systems and safety: resilience and excellence in healthcare practice. *Anaesthesia* (2018).
- [14] W. Petschnig and E. Haslinger-Baumann, Critical Incident Reporting System (CIRS): a fundamental component of risk management in health care systems to enhance patient safety. *Saf Health* **3** (2017), 27.

Impact Analysis of the Policy for Access of Administrative Data in France : A Before-After Study

Vincent LOOTEN^{a,1} and Marie SIMON^b

^aUMRS 1138, Centre de Recherche des Cordeliers, Université de Paris, Paris, France

^bUniversité Paris-Est Créteil, Créteil, France

Abstract. In 2017, French institutions reformed their data access policy regarding the national insurance and administrative databases systems (French abbreviation: SNDS), including claims data from hospitalization discharge summaries for the entire population follow-up encompassing over 10 years. Our study aimed to quantitatively and qualitatively describe such authorization before and after the reform. We extracted data access demands for French National Health Data Institute (INDS) data before and after the reforms. We included only studies that needed data extracted from the SNDS database and authorization of the regulator. We inferred the number of projects accepted pre- and post-reform, and we describe the types of studies, their topics, and the types of data used. We included 802 data access demands between January 1st 2008 and September 21st 2019. The median of data access demands by year increased from 21.5 to 203. This increase was lower in the studies included insurance data (21.5 to 70). The evolution is driven by the activity of Private companies and contract-research organization. The number of studies on Hematology and oncology and internal medicine increased respectively by 1.7 and 1.4 factors. Data access of claims data refers to the “accessible” dimension of the FAIR guiding principles. However, extrinsic factors influence the accessibility of claims data such as human factors (e.g. data scientist with experience in claims data) and economic factors (e.g. data infrastructure HIPAA and GDPR compliant).

Keywords. [Data Sharing](#), [Policy](#), [Administrative Claims](#)

1. Introduction

Data sharing is a prerequisite for open innovation in health to produce reproducible research and to promote open science. Healthcare professionals and citizens need to be equipped to share data to facilitate scientific research and minimize the potential privacy threats[1]. Since May 2018, data sharing, including that of European citizens’ personal data, as well as data processing in the European Union, has been subject to the General Data Protection Regulation (GDPR)[2]. The GDPR provides a legal framework for data reuse and offers a chance to standardize data protection practices in research, as well as opportunities for researchers in medical informatics to develop new models for information technology infrastructure[3]. In 2017, French institutions reformed their data access policy regarding the national insurance and administrative databases systems

¹ Corresponding Author: Dr Vincent Looten, Université de Paris, Paris, France; E-mail: lootenv@gmail.com.

(French abbreviation: SNDS), including claims data from hospitalization discharge summaries for the entire population follow-up encompassing over 10 years. Researchers with access to the SNDS are producing studies on various topics such as patient care pathways, pharmacoeconomics, and pharmacoepidemiology[4]. The French reform for data access to SNDS aimed to open insurance data to a wider research community, including public and private research labs. The research community hypothesized that enforcement of the GDPR has impacted this reform's effects. However, to the best of our knowledge, no previous studies have described the 2017 reform's impact on data access authorization with regard to the influence of the GDPR. Our study aimed to quantitatively and qualitatively describe such authorization before and after the reform.

2. Method

Data sources. We extracted data access demands for French National Health Data Institute (INDS) data before and after the reforms. The data used in this study are available at www.github.com/vlooten/datapolicy, while more recent data can be downloaded from the INDS website (www.indsante.fr).

Inclusion criteria. We included only studies that needed data extracted from the SNDS database and authorization of the regulator. Projects that were resubmitted were counted only once (for the first submission).

Data access policy. Before the reform, a national committee evaluated submitted projects, and an approval from the French Data Protection Authority (CNIL) was required (i.e., two authorizations were required). After the reform, all projects following the CNIL national guidelines required just one authorization from a national committee. This committee checked concordance between objectives and methods, and compatibility of the study objectives with the public interest.

Outcomes and definitions. We inferred the number of projects accepted pre- (before August 1st 2018) and post-reform (after August 1st 2018), and we describe the type of investigator, the types of studies, their topics, and the types of data used. The type of investigators described were: (1) Government and agencies (2) Private companies and Contract-research organizations (3) Hospitals and care structures (4) Citizens and associations (5) Schools and research centers. The categories of the type of the studies were: (1) Descriptive and transversal studies (2) Longitudinal or prognosis studies (3) Pharmacoepidemiology, pharmacoeconomic studies (4) Economics studies (5) Other (included methodological and unclassified studies). The topics described were: (1) Cardiology and vascular medicine (2) Hematology and oncology (3) Internal medicine (4) Psychiatry and addictology (5) Public health and occupational medicine (6) Surgery (7) Primary cares and (8) Other (included unclassified and unknown studies). The types of data described were: National data with or without insurance data.

Statistical analyses. Data were expressed as number (%). Chi2 tests (for categorical data) was used to compare groups. All tests involved use of R 3.6.1(R Foundation, Vienna, Austria).

3. Results

We reviewed 2031 data access demands between January 1st 2008 and September 21st 2019. We included 802 data access demands according to the inclusion criteria. Reasons

of exclusions were: no pairing with national administrative data (N=1146) data access rejected or canceled studies (N=41), renew data access authorization (N=32) and missing data (N=10). Years of data access demands were: 2019 (N=203), 2018 (N=257), 2017 (N=74), 2016 (N=73), 2015 (N=61), 2014 (N=51), 2013 (N=24), 2012 (N=17), 2011 (N=19), 2010 (N=15), 2009(N=6) and 2008 (N=2). The median number of data access demands by year increase from 21.5 to 203. Table 1 presents the characteristics of the data access demands. We observed a difference between the two groups for the type of investigator. The number of access of private companies increases from 26 (8.87%) to 227 (44.6%), and school and research centers decreased from 98 (33.4%) to 59 (11.6%). We observed a change in the type of the studies with a reduction of Pharmacoepidemiology, pharmaco-economic studies (from 100 (34.1%) to 99 (19.4%). The structure of the topics also changes. In particular, the number of Hematology and oncology studies increased from 27 (9.22%) to 81 (15.9%).

Table 1. Description of the studies before and after the reform

	Before the reform (N=293)	After the reform (N=509)	p value
Insurance data	293	343	<0.001
Investigator:	293 (100%)	343 (67.4%)	<0.001
Government and Agencies	12 (4.10%)	27 (5.30%)	
Private companies and CRO	26 (8.87%)	227 (44.6%)	
Hospitals and care structures	125 (42.7%)	170 (33.4%)	
Citizens and associations	32 (10.9%)	26 (5.11%)	
Schools and research centers	98 (33.4%)	59 (11.6%)	
Type of study:			<0.001
Descriptive and transversal studies	48 (16.4%)	91 (17.9%)	
Economics studies	22 (7.51%)	47 (9.23%)	
Longitudinal or prognosis studies	116 (39.6%)	254 (49.9%)	
Pharmacoepidemiology, pharmaco-economic studies	100 (34.1%)	99 (19.4%)	
Unknown	7 (2.39%)	18 (3.54%)	
Topics:			<0.001
Cardiology and vascular medicine	29 (9.90%)	51 (10.0%)	
Hematology and oncology	27 (9.22%)	81 (15.9%)	
Internal medicine	69 (23.5%)	164 (32.2%)	
Public health and occupational medicine	66 (22.5%)	102 (20.0%)	
Psychiatry and addictology	36 (12.3%)	50 (9.82%)	
Surgery			
Primary cares	5 (0.98%)	12 (4.10%)	
Other	11 (3.75%)	20 (3.93%)	

Over the 802 studies included, only 636 required insurance data. In this subgroup, the median of data access demands by year increased from 21.5 to 70. Table 2 presents the characteristics of the data access demands of this subgroup. The difference observed between the two groups are similar.

Table 2. Description of the studies included insurance data before and after the reform.

	Before the reform (N=293)	After the reform (N=343)	p value
Investigator:			<0.001
Government and Agencies	12 (4.10%)	24 (7.00%)	
Private companies and CRO	26 (8.87%)	129 (37.6%)	
Hospitals and care structures	125 (42.7%)	124 (36.2%)	
Citizens and associations	32 (10.9%)	17 (4.96%)	
Schools and research centers	98 (33.4%)	49 (14.3%)	
Type of study:			0.003
Descriptive and transversal studies	48 (16.4%)	50 (14.6%)	
Economics studies	22 (7.51%)	38 (11.1%)	
Longitudinal or prognosis studies	116 (39.6%)	172 (50.1%)	
Pharmacoepidemiology, pharmaco-economic studies	100 (34.1%)	73 (21.3%)	
Unknown	7 (2.39%)	10 (2.92%)	
Topics:			<0.001
Cardiology and vascular medicine	29 (9.90%)	31 (9.04%)	
Hematology and oncology	27 (9.22%)	56 (16.3%)	
Internal medicine	69 (23.5%)	122 (35.6%)	
Public health and occupational medicine	66 (22.5%)	61 (17.8%)	
Psychiatry and addictology	36 (12.3%)	34 (9.91%)	
Surgery	43 (14.7%)	25 (7.29%)	
Primary cares	12 (4.10%)	4 (1.17%)	
Other	11 (3.75%)	10 (2.92%)	

4. Discussion

Main results. To the best of our knowledge, we are the first to describe the effect of the French reform on the data access demands. After the reform, the median of data access demands by year increased from 21.5 to 203. This increase was lower in the studies included insurance data (21.5 to 70). The evolution is driven by the activity of private companies and contract-research organization. The number of studies on Hematology and oncology and internal medicine increased respectively by 1.7 and 1.4 factors.

Technical significance. According to the FAIR guiding principles[5] data must be Findable, Accessible, Interoperable and Reusable. Claims data have important intrinsic

characteristics in the FAIR perspective. Data access of claims data refers to the “accessible” dimension of the FAIR guiding principles. However, extrinsic factors influence the accessibility of claims data such as human factors (e.g. data scientist with experience in claims data) and economic factors (e.g. data infrastructure HIPAA and GDPR compliant). Two reasons could explain the most important demands of private operators: (1) the reform has reduced the time of treatment (only one committee compounded by specialists of data) (2) and the legal framework has been simplified ensuring a good visibility for elaborate business plans. However, the activity of the public sector has increased fewer; this could be explained by the absence of new data infrastructure and new funding strategy (extrinsic factors). In our perspective, the efficiency of a data access policy should be quantitatively and qualitatively monitored to understand the potential leverages of data reuse.

Remaining challenges. Schmidt *et al.* [6] have described the natural history of the data reuse in the Danish National Patient Registry. However, in the French context, data access demands database is not linked to scientific publications. A perspective of our work is linking data access demands and publication to describe the natural history of the French administrative data reuse. Data reuse of claim database has a natural history and this history is influenced by the data access policy.

References

- [1] S. Patil, H. Lu, C.L. Saunders, D. Potoglou, and N. Robinson, Public preferences for electronic health data storage, access, and sharing — evidence from a pan-European survey, *J. Am. Med. Informatics Assoc.* **23** (2016) 1096–1106. doi:10.1093/jamia/ocw012.
- [2] J. Demotes-Mainard, C. Cornu, A. Guérin, P.-H. Bertoye, R. Boidin, S. Bureau, J.-M. Chrétien, C. Delval, D. Deplanque, C. Dubray, L. Duchossoy, V. Edel, C. Fouret, A. Galaup, F. Lesaulnier, M. Matei, F. Naudet, V. Plattner, M. Rubio, F. Ruiz, S. Sénéchal-Cohen, T. Simon, A. Vidal, A. Viola, and M. Violleau, How the new European data protection regulation affects clinical research and recommendations?, *Therapies*. **74** (2019) 31–42. doi:10.1016/j.therap.2018.12.004.
- [3] J.M.M. Rumbold, and B. Pierscionek, The Effect of the General Data Protection Regulation on Medical Research, *J. Med. Internet Res.* **19** (2017) e47. doi:10.2196/jmir.7108.
- [4] P. Tuppin, J. Rudant, P. Constantinou, C. Gastaldi-Ménager, A. Rachas, L. de Roquefeuil, G. Maura, H. Caillol, A. Tajahmady, J. Coste, C. Gissot, A. Weill, and A. Fagot-Campagna, Value of a national administrative database to guide public decisions: From the système national d’information interrégimes de l’Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France, *Rev. Epidemiol. Sante Publique*. (2017). doi:10.1016/j.respe.2017.05.004.
- [5] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A. ‘t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*. **3** (2016) 160018. doi:10.1038/sdata.2016.18.
- [6] M. Schmidt, S.A.J. Schmidt, J.L. Sandegaard, V. Ehrenstein, L. Pedersen, and H.T. Sørensen, The Danish National Patient Registry: a review of content, data quality, and research potential, *Clin. Epidemiol.* (2015) 449. doi:10.2147/CLEP.S91125.

Legal and Ethical Issues in Secondary Use of Administrative Health Data: The Case of Latvian Healthcare Monitoring Datalink

Signe MEZINSKA^a, Arnis BUKA^b, Agnese BANKAVA^b, Juris BARZDINS^{a1}

^aUniversity of Latvia, Faculty of Medicine

^bUniversity of Latvia, Faculty of Law

Abstract. The paper presents analysis of the legal and ethical issues surrounding establishment of the Latvian Healthcare Monitoring Datalink. The paper covers three interconnected issues in the context of the use of administrative health data for research purposes – anonymization of data, concept of ‘public interest’ and involvement of research ethics committees. The analysis has been put into broader context of interaction between General Data Protection Regulation (GDPR), national legislative measures and practical needs of researchers. Neither GDPR, nor Latvian legal framework regulate the particularities on the use of potentially identifiable health data in research. Also, the practical use of ‘public interest’ as a basis for lawful processing of personal data concerning health for research purposes is not clear. More extended involvement of research ethics committees might serve as useful tool for determination the ‘public interest’ and for the evaluation of proportionality when balancing the aims of the research and the personal data protection

Keywords. Personal data, data concerning health, secondary use, anonymization, informed consent, public interest

1. Introduction

To perform their functions, several state institutions in Latvia (the National Health Service, the Centre for Disease Prevention and Control, the State Emergency Medical Service and the Health Inspectorate) are collecting personalized patient level data. The primary use of these data does not require that the personalized data stored in one institution is linked to the data of the same person in another institution. Linking these data sets through a unique identifier for each person may ensure that the data can be used more broadly, e.g. for monitoring the quality and effectiveness of health care and for research. Joint initiative of the University of Latvia and the Ministry of Health of Latvia has resulted in development of a unique Latvian Healthcare Monitoring Datalink (Datalink) of unidentifiable persons and identifiable service providers [1].

The Datalink is controlled by the Centre for Disease Prevention and Control, and it includes healthcare process-related and outcome-related data of the population. The currently available data (2014-2018) includes information (health care provider, dates

¹ Corresponding author, Faculty of Medicine, University of Latvia, Jelgavas iela 3, LV-1004, Riga, Latvia; E-mail: juris.barzdins@lu.lv

of provided services, manipulations, primary and secondary diagnoses, costs etc.) about more than 65 million ambulatory and 3,5 million stationary treatment episodes as well as over 250 million units of medicines dispensed at pharmacies (including formulation, strength, quantity, and price). The demographic file provides information on patient's age, sex, place of residence. Additionally, the Datalink includes data from national Causes of Death Registry and from registries for particular diseases (e.g. diabetes, cancer) and conditions (e.g. trauma). The Datalink is updated once a year. To ensure protection of personal data, the direct identifiers of persons are irreversibly anonymized to exclude the possibility for either party to store and reuse the encryption key. Thus, to add data each year, the Datalink is re-created.

While the use of the Datalink for the monitoring of healthcare system performance by governmental institutions is regulated by the Cabinet decree assigning this function to the Centre for Disease Prevention and Control, the use of the Datalink for research is not clearly regulated by the existing regulatory framework leading to several legal and ethical problems. The present article addresses three of these issues – anonymization of data, concept of ‘public interest’ and involvement of research ethics committees.

2. Anonymized or potentially identifiable?

The data stored in the Datalink provides opportunity for researchers to analyse highly valuable administrative health data. However, due to gaps in legal regulation and due to ethical considerations, the secondary use of such anonymized but still potentially identifiable data is complicated. Some uncertainties arise also from the fact that the data originally have been collected for administrative purposes therefore there is no informed consent obtained for secondary use of the data in research.

The main scientific value of the Datalink is a possibility to track an anonymous patient's journey through the healthcare system knowing basic information about his/her health status, healthcare interventions and outcomes. The value of research data in many cases depends on keeping details regarded also as possible indirect identifiers as patient age, sex, place of residence, service provider, diagnosis, dates and codes for interventions. At the same time, the anonymization after linking independent datasets ensure privacy and confidentiality of persons concerned only if an indirect identification is not possible. It leads to the trade-off between the public interest justifying procession of such data for research and the small, but still existing risk of identification of persons.

Some, but not all aspects of this problem are regulated by the General Data Protection Regulation (GDPR) which does not apply to fully anonymized data concerning health but applies to pseudonymized and potentially identifiable data. Article 9 of the GDPR includes data concerning health into “*special categories of personal data*” for which there is a presumption that its processing is prohibited, at the same time, part 2 of the Article 9 provides a long list of exceptions on when the processing of special categories of personal data is permitted [2]. Mostly those exceptions are related to the consent of data subject or to the overriding public interest justifying procession of special categories of personal data. However, also in these exceptional cases, data processing must be legitimized by EU or national legislation and must be proportionate, including the duty of the data processor to take “*suitable and specific measures to safeguard the rights and freedoms of the data subject*” [2]. Article 9 (2) (j) of the GDPR includes a reference to the Article 89 explaining further

how to apply the above-mentioned conditions to processing special categories of personal data. Among other, Article 89 refers to the principle of data minimization explained in the Article 5, namely, that the data must be adequate, relevant and only include what is necessary for the purposes of the processing. The preamble of the GDPR also clarifies the processing of data concerning health for research purposes in several recitals. Recital 56 of the preamble refers to the main objective that justifies the processing of special categories of personal data, including data concerning health – it must be in the public interest. Recital 159 states that the processing of personal data for scientific research purposes should be interpreted in a broad manner including, for example, technological development and demonstration, fundamental research, applied research and privately funded research. While outlining the basic principles of data protection, the GDPR at various places points out to the responsibility of Member States to adopt further legislation, providing “*an unusually wide margin of maneuver for Member States*” [3].

In 2018 Latvia adopted the Personal Data Processing Law. Article 2 of this law puts forward the aim of the law “*to create legal preconditions for the establishment of the system of protection of personal data at the national level*” [4]. Regrettably, the new law is overly brief regarding regulation of scientific research. Although the law contains Article 31 on the processing of data for scientific or historical research purposes, in substance this article is just a blanket norm referring to the GDPR and copying Article 89 (2) of the GDPR.

Neither GDPR, nor Latvian legal framework directly regulates the use of anonymous or potentially identifiable health data in research. However, both the conditions for processing fully anonymous data and data anonymized but still potentially identifiable applies to the research use of the Datalink data.

3. Public interest as a basis for secondary use of administrative health data in research

One of the possible solutions is to use public interest as a basis for secondary use of administrative health data in research. The Article 9(2) of GDPR in listing possible grounds for processing of personal data not only includes a point (j) with specific reference to the “*scientific or historical research purposes*”, but also more general point (g) that allows processing of personal data “*for reasons of substantial public interest*”. The interrelation between those two grounds in the general scheme of the GDPR is rather unclear, especially considering recital 159 which states that “*scientific research purposes should also include studies conducted in the public interest in the area of public health*”. Yet, in any case, it should be possible to use public interest as a basis for secondary use of administrative health data in research, if all the preconditions on the proportionality, mentioned in the Article 9(2)(g) of the GDPR are fulfilled.

The problematical aspects of seeking the balance between public interest in research and privacy rights of individuals have already been analysed by several authors [5]. There are particular problems regarding application of proportionality test in cases when public interest serves as a basis for conducting the research, for example: how to define and assess public interest in the context of particular research protocol?

Assessment of the social value of the research study before it has been started is one of the main requirements in international documents defining the principles of research ethics for biomedical research involving human subjects. Human subject

research as defined in these documents is not only research where a person is directly involved in a study (e.g. clinical studies of medicinal products), but also research studies on identifiable human biological material and personal data.

Article 16 (iii) of the Council of Europe Convention on the Protection of Human Rights and Dignity in Biology and Medicine states that studies involving human subjects may be carried out only if *“the research project has been approved by the competent body after independent examination of its scientific merit, including assessment of the importance of the aim of the research, and multidisciplinary review of its ethical acceptability”* [6]. A detailed requirement to assess the social value of the study is included in the International Ethical Guidelines for Health-related Research Involving Humans issued by the Council for International Organizations of Medical Sciences (CIOMS) (2016). It states that *“[t]he ethical justification for undertaking health-related research involving humans is its scientific and social value: the prospect of generating the knowledge and the means necessary to protect and promote people’s health”* [7]. In Latvia, the Law on the Rights of Patients (Section 9, paragraph 8) allows research use of patient data recorded in medical documents without informed consent if several requirements are met, including the requirement that *“the study is carried out in the public interest”* [8].

At the same time, several questions arise in the context of these requirements, like: who and how should assess the social value of a research study? what exactly makes the study socially valuable? The practice shows that there is a need for criteria for assessing the social value of a study that could be used by research ethics committees and other bodies involved in the evaluation process. CIOMS guidelines state that *“The scientific and social value of research can be difficult to quantify, but it is generally grounded in three factors: the quality of the information to be produced, its relevance to significant health problems, and its contribution to the creation or evaluation of interventions, policies, or practices that promote individual or public health”* [7]. Ethicist Ezekiel Emanuel and colleagues have defined the social value of the study as the instrumental value of the new knowledge for improving human health. They clarify that the two main reasons for assessing the social value of a study by ethics committees are the prevention of harming study participants (including breaches of privacy) and the waste of limited research resources [9]. The accessibility and dissemination of the study results, including publication of negative results also is an essential pre-condition for ensuring the social value of the study.

4. New role for research ethics committees?

The GDPR in the Article 40 as well as in the Recital 98 encourages *“the drawing up of codes of conduct intended to contribute to the proper application of this Regulation”*. It leads to the role that research ethics committees might play in evaluation of public interest and protection of research participants. Currently in Latvia, the ethics approval for research protocols is clearly required by law only in few specific cases: clinical trials of medicinal products, clinical trials of medical devices and human genome research. For clinical trials and studies using data from medical documents mentioned in the Law on Patient Rights there is no legal requirement for ethics committee review.

This demonstrates the need for re-definition of the role of ethics committees in Latvia. The scope of research requiring review by research ethics committee must be broadened, because the current approach does not meet the requirements of

international documents on research ethics, as well as it does not ensure assessing of public benefit and protection of research participants for all types of medical and health research.

5. Conclusions

There are several ambiguities and problems in the context of establishing and use of the Datalink for the research purposes, which are not fully addressed by the existing legal framework in Latvia. First, it is not completely clear how to define and evaluate the public interest in the context of health research. Second, the line between fully anonymized and potentially identifiable data, as well as evaluation of risk of identification needs further clarifications. Third, there is a need to increase a role and involvement of research ethics committees in the review of secondary use of data concerning health.

The GDPR was one of the most important and most massive legal developments in the EU law of 21st century, ambiguities in the context of research in the field of data concerning health being only one small fraction of challenges that follows the GDPR. One can only hope that future case-law of the Court of Justice of the EU as well as additional national legislative measures will help to deal with those challenges.

Acknowledgement

This study is a part of the UL/CDPC joint project “Transparency and healthcare data - towards public monitoring for quality and efficiency”.

References

- [1] University of Latvia & Centre for Disease Prevention and Control, *Health Personalised Data Guidebook for Health Care Monitoring Datalink*, 2019. Retrieved October 1, 2019, from http://med.oranzais.lumii.lv/in_english.html
- [2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016
- [3] V. Chico, The impact of the General Data Protection Regulation on health research, *British Medical Bulletin*, 128(1) (2018), 109-118.
- [4] Republic of Latvia, Personal Data Processing Law, 2018. Retrieved October 1, 2019, from <https://likumi.lv/ta/en/en/id/300099-personal-data-processing-law>
- [5] J. Bell, S. Aidinlis, H. Smith, M. Mourby, H. Gowans, S. Wallace, J. Kaye, Balancing Data Subjects' Rights and Public Interest Research: Examining the Interplay between UK Law, EU Human Rights Law and the GDPR, *European Data Protection Law Review* 5(1) (2019), 43-53.
- [6] Council of Europe, *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine* (2019)
- [7] CIOMS, *International Ethical Guidelines for Biomedical Research Involving Human Subjects* (2016)
- [8] Republic of Latvia, Law on the Rights of Patients, 2009. Retrieved October 1, 2019, from <https://likumi.lv/ta/en/en/id/203008-law-on-the-rights-of-patients>
- [9] E.J. Emanuel, D. Wendler, Ch. Grady, An Ethical Framework for Biomedical Research. In *The Oxford Textbook of Clinical Research Ethics*. Edited by E.J. Emanuel and Ch. Grady. Oxford University Press. New York, 2008.

The Definition of Informatics Competencies in Finnish Healthcare and Social Welfare Education

Alpo VÄRRI^{a,1}, Minna TIAINEN^b, Elina RAJALAHTI^c, Ulla-Mari KINNUNEN^d,
Lea SAARNI^b and Outi AHONEN^c

^a *Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland,*

^b *Tampere University of Applied Sciences, Tampere, Finland*

^c *Laurea University of Applied Sciences, Espoo, Finland,*

^d *Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland*

Abstract. Finland is a world leader in the use of public electronic services. Continuous improvement to competencies is a prerequisite for the success of digitalisation in the service development sector. The increasing use of information technology in health and social care needs to be taken into account in the education of the health and social care sector work force. The mandate of the national SotePeda 24/7 project is to identify and define the informatics competencies required for multidisciplinary education of this sector in Finland. The project has adapted international recommendations for use in the national context. The national recommendation covers 12 areas of competency and related content. In addition to defining competencies, the project has produced a toolbox of materials for use by educators of these topics in universities that cover applied sciences and lifelong learning. The results of the project are expected to significantly improve the preparedness of graduating health and social care and related engineering and business sector students to make full use information technology, all of which benefits the national health and social welfare system.

Keywords. competence, skill, informatics, information technology, health care, social care, education

1. Introduction

Finland has the highest percentage of eHealth service users in the European Union (EU), (i.e., representing almost 50% of users). Generally, only 18% of EU citizens use online health care and social welfare² services [1]. A key objective of the National Finnish eHealth and eSocial Strategy [2] is to support the renewal of the healthcare and social welfare sectors. It emphasises the active role of citizens in maintaining their own well-being, which can be achieved by improving information management and increasing the

1 Corresponding Author, Tampere University, Hervanta Campus, Sähkötalo, FI-33014 Tampere University, Finland, Alpo.Varri@tuni.fi.

2 In Finland, 'social welfare' is social services, social assistance, social allowance, social loans and related measures intended to promote and maintain social security and the functional capacity of individuals, family and the community.

provision of online services; in this regard, the role of healthcare and social welfare professionals is crucial. A primary goal is that, by 2020, professionals in health care and social welfare will have access to information systems that support their work and its operating processes. The measures extend to national responsibility, for example, of training professionals in information management, as well as data entry, data protection, information security and knowledge management. The eHealth strategy of the Finnish Nursing Association 2015–2020 incorporates these objectives, with the aim of imparting competencies to nursing students in the use of technology, information literacy, knowledge management and informatics processes. These competencies will also be updated with supplementary training [3].

The increasingly widespread use of information technology [4] to support healthcare and social welfare professionals has created a need to improve informatics competencies in Finland. The Ministry of Education and Culture in Finland financed a national project, the Multidisciplinary Competencies for Health Care and Social Welfare Services in Finland (SotePeda 24/7) [5], with the primary aim of defining necessary informatics competencies and producing educational material for multidisciplinary use at the bachelor level education in the fields of health and social care and related engineering and business sectors of universities of applied sciences (UAS) in Finland. A secondary aim of the project was to support education that prepares students for work in a multidisciplinary team with a view to promoting the health and well-being of customers with the help of information technology. This paper presents how the competence definition work was carried out and what its present result is.

2. Method

The knowledge base for the descriptions of the competencies was primarily founded on international definitions of health informatics competencies in the Health Information Technology Competencies (HITComp) database [6,7]. The second version of the HITComp database which we used, is a result of the EU-US eHealth Work Project, an EU-funded project [13]. It was designed for educators, workforce developers, current and future workforce members, students, eHealth managers, staffing authorities, and others in healthcare information technology or eHealth [7]. Our project team used HITComp to search for competencies in a variety of healthcare roles. Competencies (i.e., ‘direct patient care’ and ‘administration’) were sorted into domains. The ‘basic’ and ‘baseline’ level competencies of HITComp were included for further elaboration. A healthcare professional with good knowledge of Finnish and English translated the competencies into Finnish and a group of health informatics professionals checked the correctness of the translations.

To adapt the list of competencies to the national context, the descriptions of the competencies also took into account the competence need findings of the latest studies in Finland based on a literature search. These included the Finnish Care Classification System [8], the eHealth Strategy of the Finnish Nurses Association 2015–2020 [3], the Nurses’ Competence Survey [9], national eHealth and eWelfare system (i.e., Kanta) documentation [10] and research into the competencies of healthcare professionals [11, 12]. The project team of around ten people, consisting mainly of experts in nursing but also participants with social welfare, health informatics and biomedical engineering backgrounds discussed the composition of competencies until a consensus was reached.

The HITComp originated descriptions of the competencies were grouped in the same way as those for the EU-US eHealth Work Project [13]. Areas of competence were initially described using a mind map, which made them clearer and easier to read. Finally, the mind-map was converted to a tabular form which then became the version 1.0.

3. Results

The produced list of informatics competencies required for healthcare and social welfare professional education in Finland can be seen in Table 1. HITComp terminology was used or adapted for the definitions.

4. Discussion

Although international recommendations for informatics competencies for health care professionals are available [6, 7, 14, 15], they are only applicable to a certain extent in Finland. A need remains to adapt the curricula to include components that support the national health care and social welfare environment and culture, as well as national strategic objectives [2, 3]. The competency definitions produced by the SotePeda 24/7 project [5] cater to these needs. For example, the proposed recommendations include social welfare-related competencies that are frequently absent from international recommendations, in addition to knowledge relating to the national health and welfare information repository, Kanta [10]. These recommendations also align with the Finnish eHealth and eSocial Strategy [2], which recognises that work is performed in multidisciplinary teams and sometimes includes business or technology professionals.

The SotePeda 24/7 project will reach completion at the end of 2020; until then, the work on definitions of informatics competencies continues. The 12 defined areas of competence will remain in the final version of competency recommendations. However, any additional feedback obtained from interactions with social welfare and health care authorities may necessitate minor improvements having to be made to the recommendations.

The proposed descriptions of the informatics competencies were valid at the time of writing. Digitalisation, innovations in information technology and new system implementations [4] will, however, necessitate updates to the descriptions from time to time. Each country should have a process in which competent professionals update the competence definitions. Finnish educational authorities on healthcare and social welfare informatics produced the recommendations, and the results will be followed and commented by authorities working in key healthcare and social welfare service positions and research institutions.

In addition to the competency definitions, the SotePeda 24/7 project is also responsible for the production of educational materials that support the educational approaches relating to these competencies. For example, one work package elaborates on educational materials about the ethical aspects of care digitalisation. Specifically, a toolbox has been developed for use by UAS educators (but potentially in other levels, too), who can select the materials and method of instruction of choice. The intention is

that the materials should support monological, dialogical and triological learning. The materials will be stored in a national repository of educational material.

Table 1. Definitions of health care and social welfare informatics competencies.

Area of competence	Main content
Basic information and communications technology (ICT) competencies	Information management and its tools, information and communication technologies and information literacy skills. Introducing new operating models, anticipating service and training needs, anticipating future digital literacy skills and motivation.
Online interactive competencies	Online dialogue with persons, changes in the roles of professionals, electronic communication environments, online meetings and consultations and the use of social media and online services.
Service competencies in digital health and the social care sector	Health and social care service structures, the usefulness of electronic health services, different electronic service environments and tools, the roles of social and healthcare actors, electronic service pathways, e-services and virtual reception.
Person-centred guiding competencies in a digital environment	Assessing customers' IT skills, IT guidance for customers, directing customers in search of information, supporting self-care by clients, the preparation of electronic guidelines, the production of online material, electronic outpatient clinics and information services, ie. chat rooms.
Competencies to monitor health and well-being in a digital environment	Artificial intelligence, sensors, robotics, wearable technology, utilities, various monitoring tests and instruments, monitoring information literacy and assessing the reliability and adequacy of information.
Health and social care informatics competencies	The interoperability of electronic systems, the communication of information via electronic information systems, electronic logging, national patient and social welfare data repository Kanta.fi, electronic data storage and roles and responsibilities regarding the use of information and legislation.
Multi-stakeholder service co-development competencies	Assessing the reliability of data sources, information management guidance and collaboration, privacy and security, co-operation and new operating models.
Ethical competencies	Ethical operating models and ethical competence in electronic services.
Service design competencies	User orientation, participation, innovativeness and new service pathways.
Knowledge management competencies	The use of monitoring and research data, customer- and patient-specific information, availability, quality and effectiveness of services (e.g., considering changing needs).
Research, development and innovation competencies	Assessment and continuous improvement of one's own skills, work community skills development, the development of electronic services, quality criteria for electronic services, the development of health and well-being technologies, exploitation of evidence-based information and an evaluation of effectiveness.
Societal competencies	Continuous consideration of information security in operations, the social impact of health technology on well-being and daily life, digital democracy and the promotion of social inclusion.

5. Conclusion

This project has significant national responsibility as it comprises recommendations on the healthcare and social welfare informatics educational material and curricula content of nursing students, other healthcare and social welfare professionals, as well as those in related fields (i.e., engineering and business) in UASs in Finland. As almost all UASs that offer this type of education in Finland are involved in the project, it is likely that the

recommendations will be followed at least to some extent. The educational material produced by the project will assist educators to introduce these topics to their students. The present version of the recommendations defines 12 areas of competence, the corresponding content of which is defined in detail. The recommendations are supported by a toolbox of educational materials to be used freely by educators in relevant fields in Finnish universities. It is expected that the recommendations and online educational materials will significantly improve basic healthcare- and social welfare informatics-related education in Finland.

References

- [1] European Commission 2018, *Digital Economy and Society Index Report 2018*. Digital Public Services. Available at: <https://ec.europa.eu/digital-single-market/en/digital-public-services-scoreboard>. Accessed September 13, 2019.
- [2] Ministry of Social Affairs and Health, *Information to support well-being and service renewal. eHealth and eSocial Strategy 2020*. 2015.
- [3] O. Ahonen, P. Kouri, P. Liljamo, H. Granqvist, K. Junttila, U-M. Kinnunen, et al., *The eHealth Strategy of Finnish Nursing Association 2015–2020*, 2015. Available at: https://nurses-fi-bin.directo.fi/@Bin/70790ee5a1d6ad5cfa6a57606000ecb/1506020626/application/pdf/237208/eHealth_RAPORTTI%20_ENGLANTI.pdf.
- [4] T. Vehko, S. Ruotsalainen, H. Hyppönen (eds.), *E-health and e-welfare of Finland. Checkpoint 2018*, National Institute for Health and Welfare. 2019. Available at http://www.julkari.fi/bitstream/handle/10024/138244/RAP2019_7_e-health_and_e-welfare_web_4.pdf?sequence=4&isAllowed=y.
- [5] A. Värri, U-M Kinnunen, P. Pöyry-Lassila, O. Ahonen, The national SotePeda 24/7 project develops future professional competencies for the digital health and social care sector in Finland, *Finnish Journal of eHealth and eWelfare* **11** (2019), 232-235.
- [6] A. Värri, R. Blake, J. Roberts, S. Fenton, M. Cleary, S. Zacks, G. Datta, R. Kaye, J. Parker, C. Nguyen, M.F. Cunningham, Transatlantic collection of health informatics competencies, *Finnish Journal of eHealth and eWelfare* **8** (2016), 127-136.
- [7] *Health information technology competencies database*. 2018. Available at <http://hitcomp.org/> Accessed September 11, 2019.
- [8] U-M. Kinnunen, et al. *FinCC classification system, user's guide. FinCC 4.0*, National Institute for Health and Welfare, Helsinki, Finland, 2019, <https://yhteistyotilat.fi/wiki08/display/FLKJ1>.
- [9] U-M. Kinnunen, T. Heponiemi, E. Rajalahti, O. Ahonen, T. Korhonen, H. Hyppönen. Factors related to health informatics competencies for nurses-results of a national electronic health record survey, *Computers, Informatics, Nursing* **37** (2019), 420-429.
- [10] *The Finnish National eHealth and eWelfare System Kanta*. 2019 Available at: <https://www.kanta.fi/en/what-are-kanta-services>.
- [11] E. Rajalahti. *The development of health educators' nursing informatics competence* (Finnish). 2014. Available at: http://epublications.uef.fi/pub/urn_isbn_978-952-61-1611-2/
- [12] S. Kujala, E. Rajalahti, T. Heponiemi, P. Hilama. Health professionals' expanding eHealth competencies for supporting patients' self-management, *Studies in Health Technology and Informatics* **247** (2018), 181-185. doi: 10.3233/978-1-61499-852-5-181
- [13] *EU-US eHealth Work Project*. Available at: <http://www.ehealthwork.eu/>.
- [14] J. Mantas, E. Ammenwerth, G. Demiris, A. Hasman, R. Haux, W. Hersh et al., IMIA recommendations on Education Task Force. 2010. Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health informatics. First Revision, *Methods Inf. Med* **49** (2010), 105-120.
- [15] J. Sensmeier, C. Anderson, T. Shaw, International evolution of TIGER informatics competencies, *Stud Health Technol Inform*, 232, 69-76.

This page intentionally left blank

Section 7

Swiss Personalized Health Network

This page intentionally left blank

SPHN - Development and Usability Testing of an Electronic General Consent Prototype

Julia MAURER^{a,1}, Ramon SACCILOTTO^a, Jörg WILLERS^a, Pascal DÜBLIN^a, Vu DUONG^a and Christiane PAULI-MAGNUS^a

^aDepartment of Clinical Research, University and University Hospital Basel, Basel

Abstract. Background According to the Swiss Law on Research in Humans, the reuse of routinely collected genetic and non-genetic data and samples from patients for research purposes requires the consent of patients. Unfortunately, the so far established paper-based processes are intrinsically linked to the hospital admission process, labour intensive and not yielding the targeted return rates. Therefore, the overall goal of the presented SPHN project is to increase patient reach by providing hospitals with a patient-centric, user-friendly and admission-independent electronic general consent pathway. As part of the project, feasibility of different digital pathways was evaluated in a usability testing.

Methods Based on a nationwide harmonised template, a mobile centric progressive web application was developed by the Department of Clinical Research Basel. Usability of the application and according user journeys were evaluated at all partner hospitals. Two options of giving consent were explored using 1) patients' smartphones without any involvement of hospital personnel and 2) a hospital device (tablet) with explicit confirmation of patient identity by hospital personnel. Participant signatures were captured as a picture of a handwritten signature on paper taken with the camera of the smartphone or tablet. Usability issues and feedback of participants were documented directly after usability testing.

Results In total, 122 users agreed to participate in the usability testing using a tablet or smartphone. The general consent request workflow on the smartphone or tablet was regarded as user friendly and easy to navigate by 96% of all participants. However, capturing a picture of a handwritten signature resulted in usability issues in multiple cases, i.e. due to missing pen or paper.

Conclusion Usability testing of our prototype application showed a broad acceptance of participants regarding the use of mobile electronic devices to give general consent. Therefore, we believe that easy-to-use digital general consent processes provide effective means to increase the patient pool for health-related research. Further discussions with legislative bodies are required to find patient centric, feasible and legally acceptable solutions in the specific case of electronic general consent for the near future.

Keywords. SPHN, personalized health, electronic general consent

1. Introduction

In Switzerland, use of routinely collected personal data (genetic and non-genetic) and biological material for research requires written consent of the affected person or the

¹ Corresponding Author, Julia Maurer, Department Clinical Research, Universitätsspital Basel, Schanzenstrasse 55, 4031 Basel; E-mail: julia.maurer@usb.ch

respective legal representative according to the Human Research Act (HRA) in January 2014 [1]. Today, all Swiss university hospitals have implemented rules and procedures to assure information of patients and documentation of consent in accordance with HRA requirements (so called «general consent process»).

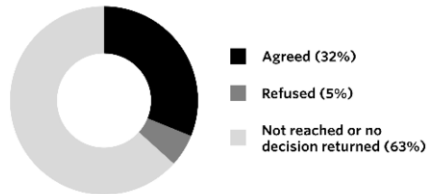


Figure 1. Consent data from 2018 at the outpatient and inpatient sector of the University Hospital Basel.

Presenting paper-based consent forms to patients and collecting responses is labour intensive and comprises several individual process steps. Analysis of consent data from 2018 at the University Hospital Basel revealed that 36.7 % of patients in the outpatient and inpatient sector (47'290 out of 128'883) returned a decision. Of those patients, 86.1 % (40'718 out of 47'290) agreed to share their data and samples. As shown in Figure 1, 63.3 % of patients at the hospital where either not reached or did not return the general consent form. An internal survey revealed, that the situation is similar at all five Swiss university hospitals.

Predominant reasons for low return rates show the urgent need to develop additional, patient-centric and easy-to-use methods to inform patients and electronically collect general consent information. Furthermore, it is important to improve understandability of the general consent information employing visually appealing and interactive presentation that support patient understanding and satisfaction [2, 3].

Thus, the goal of the presented project was to evaluate the feasibility of an electronic general consent pathway through usability testing of a mobile prototype application.

2. Methods

A mobile progressive web application was developed by the Department of Clinical Research Basel as usability prototype based on the content of the harmonised national general consent template [4]. The aim was to find unforeseen issues with the consent workflow and user interaction when using the application in a real-world setting.

Two electronic general consent pathways were explored:

- a) Giving general consent using the personal smartphone without any involvement from hospital personnel. Identity of participants was verified by a token supplied through a mobile short message or email message for which participants provided the respective contact information themselves.
- b) Giving general consent using a hospital device (i.e. tablet) and manual confirmation of patient identity through hospital personnel.

Usability testing was performed at different settings (inpatient/outpatient), clinics and different patient population/participants (children/adults/elderly, male/female). Patients attending the hospital were asked by specifically instructed recruiters to participate in the testing. Participants using their own smartphone were provided with a language specific brochure containing a QR-code that could be scanned to directly access the mobile application website. In the case of hospital devices, hospital personnel would open the general consent website from a pre-installed app icon and hand it over to the patient. Visual features of the application were intentionally designed in close resemblance with hospital corporate design to establish a trust basis for participants. The landing page was presented in the same language as the brochure, with the possibility to choose a different language on request (DE, FR, IT, EN).

On a subsequent screen, the key points of the general consent content were summarized, followed by the full consent information as well as a hospital specific contact address for further information or questions. Participants would subsequently choose to agree or refuse the general consent and provide personal information (last name, first name and date of birth) with the option to specify a legal representative. In case of smartphone only use, identity was then confirmed via a token supplied by mobile short message or email for which participants provided the respective contact information. As a next step, participants would sign on the back of the brochure and capture the signature using the camera of the mobile device. Participants were then given the option to directly download a copy of the general consent or receive a copy via email. In case of verification through hospital personnel, participants would hand the device over to hospital personnel, to enter a secret division code, followed by manual confirmation of patient identity and manual electronic signature on the device. Evaluation was done on a pseudonymized dataset with anonymized participant comments.

3. Results

Usability testing was conducted from October 2019 until the end of 2019. All Swiss university hospitals and the University Children Hospital Basel agreed to participate in the usability testing at their institutions. A total of 146 participants were asked to participate. 122 participants (56 smartphone/66 tablet) completed the usability testing and subsequent questionnaire. The participant demographics were heterogeneous with a large age range from child to elderly. Reasons for non-participation varied from refusal and not feeling comfortable in participating to technical problems of not having a smartphone or missing internet connectivity.

In general, the electronic solution was well perceived by participants. On average, completion of the general consent process and usability questionnaire required approximately five minutes. Most participants agreed that the interface and the navigation was user-friendly. However, there were some issues regarding the signature workflow, mainly because participants did not have a pen at hand or had difficulties when taking a picture with the mobile device. In addition, participants criticized the amount of information which was provided as full consent text, «Too much to read»²,

² «Beaucoup à lire»

«Too much text»³ or «For me very comprehensible, but a lot – for people, not reading so fast, probably too much to read and too much to scroll»⁴.

Most usability testing participants were used to handle digital information on their smartphones or tablets. In addition, most participants did not have any concerns sharing personal information such as name, birthdate, phone number or email address as a mean to verify their personal identity. Perceived safety of the electronic consent process was high for both pathways (94%), but lower for smartphone users, where identification occurred via mobile short message / email verification only (89%). Respective concerns about data protection and the safety of an electronic signature in general were also raised explicitly by various participants, «Taking a picture of the signature does not appear to be safe»⁵, « I found the information that genetic analysis was being done somewhat frightening»⁶.

4. Discussion

Electronic general consent pathways build on the possibilities arising from digitalization, presenting information in more user-friendly and flexible ways.

An average completion time of five minutes and our own observations show that most users skip reading the full general consent text. This behavior is already known for the acceptance of general terms and conditions, nowadays a common pre-requisite for many online services. Feedback associated with length of text, indicated, that summary information with links to additional in-depth content would be perceived as more user-friendly and approachable. This could be even further improved by the use of interactive features such as dynamic infographics or video content, especially in the context of care for children and young adults.

Current legal requirements in Switzerland mandate a written consent for the use of data for research purposes (HRA Art. 16). Due to a missing definition of «writing» in this context, there is an ongoing legal debate whether consent has to be given in the form of a handwritten wet-ink signature or whether other means of «written consent» would be applicable. Usage of electronic identities and qualified electronic signatures would provide a legally equivalent digital process to a wet-ink signature [5]. Unfortunately, the implementation of electronic identities in Switzerland is still in its infancy and not a feasible solution in the near future.

The presented approach with a paper-signature and subsequent digitalization in the form of a photography is clearly not suited in terms of usability and legal aspects. However, general consent is used for the re-use of data and samples constituting a substantially lower risk compared to a study specific informed consent. Consequently, a less rigorous form of written consent – i.e. a handwritten signature on a mobile device – could provide a viable solution for general consent, if coupled with a reliable form of patient identification, such as verification through hospital personnel or biometric identification measures.

³ «zu viel Text»

⁴ «Für mich sehr verständlich aber viel - für Leute, die nicht so schnell lesen wohl zu viel und auch zu viel zum Scrollen»

⁵ «Das Fotografieren der Unterschrift erscheint mit nicht ganz so sicher»

⁶ «Die Information, dass genetische Analysen gemacht werden fand ich etwas beängstigend»

5. Conclusion

Further discussions with involved legislative parties are required to find legally acceptable, feasible, and easy-to-use solutions in the specific case of electronic general consent for the near future. Once the solutions are feasible at daily hospital business, we believe that a digital general consent process has the potential to further increase the pool of health-related routinely collected patient data and leftover samples available for further research projects.

Acknowledgements

This project was supported by the Swiss Personalized Health Network (SPHN) initiative (Grant number: 2017DEV04). We thank our collaborating partners Francisca Jörger (University Hospital Zurich), Alex Leichtle (University Hospital Bern), Marc Pfister (University Children's Hospital Basel), Cindy Roth (University Hospital Lausanne), Caroline Samer (University Hospital Geneva) and their teams for their contribution. We also thank Annette Magnin (Swiss Clinical Trial Organization (SCTO)), Klara Posfay-Barbe (SwissPedNet) and Christine Currat (Swiss Biobanking Platform) for their support.

References

- [1] The Federal Assembly of the Swiss Confederation, Federal Act on Research involving Human Beings, Human research Act, in CC810.30. 2012.
- [2] Grady C, Cummings SR, Rowbotham MC, McConnell MV, Ashley EA, Kang G. *N Engl J Med.* 2017 Mar 2;376(9):856-867. doi: 10.1056/NEJMr1603773.
- [3] Synnot A, Ryan R, Pictor M, Fetherstonhaugh D, Parker B *Cochrane Database Syst Rev.* 2014 May; 2014(5): CD003717. Published online 2014 May 9. doi: 10.1002/14651858.CD003717.pub3
- [4] Universitaere Medizin Schweiz. Harmonized General Consent Template. 2019; Available from: https://www.unimed.ch/download_file/view/197/377.
- [5] Bundesamt fuer Kommunikation BAKOM. Elektronische Signatur. 2016 31.1.2020; Available from: <https://www.bakom.admin.ch/bakom/de/home/digital-und-internet/>

SPHN – The Swiss Personalized Health Network Initiative

Adrien K. Lawrence^{a,1}, Liselotte Selter^a, and Urs Frey^{a,b}

^a*Swiss Personalized Health Network, Swiss Academy of Medical Sciences, Laupenstrasse 7, 3001 Bern, Switzerland.*

^b*University Children's Hospital Basel (UKBB), Spitalstrasse 33, 4056 Basel, Switzerland.*

Abstract. The Swiss Personalized Health Network (SPHN) is a national initiative designed to promote the development of personalized medicine and personalized health in Switzerland. SPHN contributes to the development, the implementation, and the validation of coordinated infrastructures in order to make health-relevant data interoperable and shareable for research in Switzerland. To this end, SPHN rallies all decision-makers from key clinical, research-, research support institutions and patient organizations around the same table.

Keywords. SPHN, personalized health

1. Introduction

Given the tremendous amount of health data available, health care and medical decisions should no longer be solely based on population averages, but take into account individual patient characteristics, including the variability in genes, molecular biomarkers, environment and lifestyle. To leverage the immense potential of these data for better prevention, improved medical practice and the development of innovative treatments, transdisciplinary scientific research, as well as specific infrastructural efforts are necessary: Health data need to be interoperable and broadly accessible for research according to the FAIR principles [1]. Additionally, research involving sensitive human data demands potent IT infrastructures that permit the collection, storage and analysis of high quality data, adequately link to biobank samples and, at the same time, fulfill stringent data protection and information security requirements. Furthermore, patients and citizens have to be included, not only for consent to release their data or samples for use in research, but also as active partners supporting research and helping to advance medical knowledge. These issues raise unprecedented ethical and legal (e.g., data protection) questions that must be considered even more within an extended personalised health ecosystem [2].

Although the public Swiss healthcare system is universal, healthcare is provided by a combination of public, subsidized private and totally private systems, and is organized largely on the level of individual cantons. To a large extent, this federalist principle until now prevented a nationwide coordination and/or harmonisation of

¹ Dr. Adrien K. Lawrence, Swiss Personalized Health Network, Swiss Academy of Sciences, Laupenstrasse 7, 3001 Bern, Switzerland; E-mail: a.lawrence@sphn.ch.

biobanks, electronic clinical information systems and clinical data management infrastructures and created a heterogeneity of health data infrastructures which has retarded the development of a nationwide personalised health ecosystem as compared to countries with more homogenous national health systems. These deficiencies have been realised early- on and appropriate initiatives have been undertaken on local and regional levels. In order to expand the local initiatives to the whole of Switzerland and complement them with nationwide efforts for making health data findable, accessible, interoperable and usable for research (FAIR) on a national scale, the Swiss Government launched the Swiss Personalized Health Network (SPHN) initiative in 2017. The mission of the initiative is to contribute to the development, the implementation, and the validation of nationally coordinated infrastructures which are required to make health-related data interoperable and shareable for research in Switzerland. Given its high quality health care network and excellence in research and technologies, Switzerland has the potential *to become one of the European leaders in health and medical science*, although the regulatory aspects of the federated health system are a major challenge for big data research. The initiative shall thus help to push Switzerland to the international forefront of personalised health-related research and health care.

2. Mandate and governance

The Swiss Personalized Health Network (SPHN) is an initiative of the Swiss federal government, namely the State Secretariat for Education, Research and Innovation (SERI) and the Federal Office of Public Health (FOPH). SPHN is led by the Swiss Academy of Medical Sciences (SAMS) in collaboration with the Swiss Institute of Bioinformatics (SIB). A total of CHF 68 million was allocated to the initiative for the first funding period 2017-2020. As defined in the corresponding mandate, the long-term goal of the SPHN initiative is to establish a Swiss network in personalized medicine, in which all relevant research institutions and hospitals in this field are involved. The development of a nationally coordinated data infrastructure including interoperability of local and regional information systems in order to optimize the use of primarily clinical but also other health-related data for research in the field of personalized medicine, is the central focus of the first funding phase. Rather than building a new centralized database, SPHN adopts a decentralized approach and aims to establish interoperability of health-related information by building a dynamic scalable network of data providers based on common standards for formats, semantics, governance, and exchange mechanisms. It is therefore our mission to lay the foundations needed to establish a nationwide exchange of health-related data for better disease prevention, improved medical practice and ground-breaking innovative treatments, transdisciplinary scientific research, as well as specific infrastructural efforts.

The SPHN project organization involves partners at national level, at technical level, and at institutional level (Figure 1). SPHN rallies all decision makers from key clinical, research and research support institutions around the same table. The combined knowhow, experience and commitment of the numerous partners and experts of the SPHN Boards and Working Groups enable discussion, collaborations and progress.

The National Steering Board (NSB) is the highest body of SPHN and is responsible for the overall strategy, for maintaining regular contacts to the superordinate political authorities and for external communications in coordination with the Swiss Academy of Medical Sciences (SAMS). The NSB includes representatives from key institutions in

Switzerland (e.g. University Hospitals, universities, ETH-Domain, swissuniversities, FOPH, SNSF, patient organisation). In addition to the NSB, SPHN is composed of the following Expert Advisory Groups/Bodies:

- ELSI Advisory Group (ELSIag).
- National Advisory Board (NAB, foreseen in 2020).
- International Advisory Board (IAB).
- Data Coordination Center (DCC) including several technical Working Groups (operated by the Personalized Health Informatics Group of SIB).

SPHN is further supported by the expert knowledge of several mandated SPHN *ad hoc* Working Groups, which are responsible for solving specific issues related to infrastructure development. SPHN has put very strong efforts into close collaboration with other stakeholders from the biomedical research landscape. This includes funders such as the Swiss National Science Foundation (SAKK) or the ETH Domain Personalized Health and Related Technologies Program (PHRT), and also research networks such as the Swiss Group for Clinical Cancer Research (SAKK), the Swiss Biobanking Platform (SBP), the Swiss Clinical Trial Organization (SCTO), as well as patient organizations (e.g. ProRaris, for rare disease). During the last years, these organizations managed to identify common issues and obstacles, which need to be addressed. Typically, the above-mentioned Working Groups are now highly coordinated within these stakeholder groups.

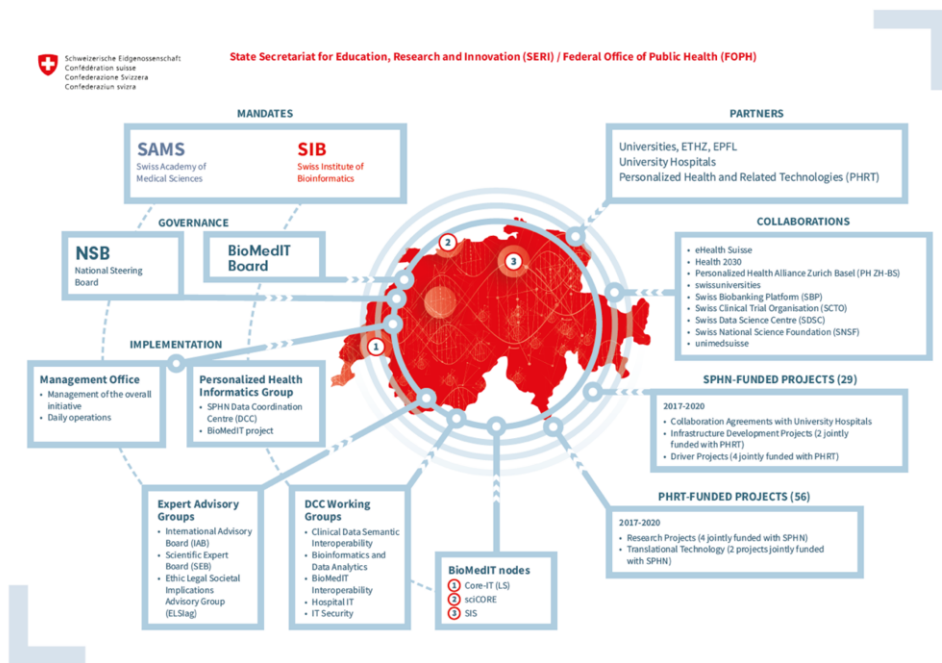


Figure 1. SPHN Governance.

3. Implementation

The SPHN initiative supports the development and implementation of coordinated infrastructures by means of a 3-pillar funding strategy:

1. **Top-down:** funding of compatible data management systems in the University Hospitals through Collaboration Agreements, as well as the SPHN Data Coordination Center (DCC) and ELSI projects.
2. **Bottom-up:** selection of projects through competitive Calls for Proposals to lead the development of infrastructures and test it with concrete research questions (Infrastructure Development Projects and Driver Projects). In 2017 and 2018, two SPHN Calls for Proposals were organized in close coordination with the Strategic Focus Area 'Personalized Health and Related Technologies' (PHRT) of the ETH Domain. Selected from a total of 76 proposals requesting CHF 90.4 million, SPHN funded 24 projects (incl. 6 co-funded by PHRT) for a total amount of CHF 25.3 million during this funding period.
3. **IT infrastructure networks:**, a secure and cutting-edge IT environment (BioMedIT, a project of SIB) is established to support computational, biomedical research and clinical bioinformatics, ensuring data privacy [3].

Further information on the projects is available on the SPHN website (www.sphn.ch).

The SPHN Management Office (MO) together with the SIB Personalized Health Informatics Group (PHI) are responsible for successfully implementing the mandate of

SPHN. The MO manages the daily operations of the initiative, coordinates the funding and ensures a good governance. The PHI group is in charge of the Data Coordination Center and the BioMedIT project.

To ensure nationwide interoperability of biomedical data, a variety of technical, ethical and legal issues have to be tackled. The ELSIag is responsible for the ethical and legal questions related to SPHN activities. The technical aspects related to the development of coordinated infrastructures, compatible data management systems, interoperability of data, and governance of guidelines, are tackled by the Working Groups of the Data Coordination Center in collaboration with the University Hospital IT teams.

4. Conclusion and Outlook

The SPHN funded projects involve 35 Swiss organizations and institutions which demonstrate the magnitude of this national endeavor. Over the past years, we have identified several key success factors for the development of a nationwide network such as i) the need for a common understanding of the vision and scope; ii) commitment of all players to share health data for the benefit of society (citizens, patients, hospitals, research institutes, etc.); iii) a transdisciplinary approach between clinicians, researchers, bioinformaticians, machine learning experts, etc.; iv) nationwide interoperability of health data and nationally coordinated data infrastructures; v) process innovation in research and health care; vi) dialogue with citizens and patient groups; and vii) transparent communication concepts to enhance public trust.

A review of the overall progress of the initiative and gap analysis was performed in 2019 based on the first project progress reports and an external international review by the SPHN International Advisory Board (IAB). As a result from this first evaluation, the biggest achievement of SPHN during the past years is probably the increasing awareness of the systemic gaps and problems related to the field. Though the initiative is overall on a good track and has made significant progress, the gap analysis revealed significant obstacles and hurdles. Most of them related to the federated health care system and the strong local awareness and particularities of data protection issues as well as ethical and legal priorities. In order to achieve interoperability, these issues, priorities and policies need to be coordinated on a national level. Here, the willingness of the stakeholders to agree and implement common strategies, standards and guidelines is a prerequisite. The political discussion needs to be initiated whether or not this should be enforced.

References

- [1] M. Wilkinson, M. Dumontier, I. Aalbersberg, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3** (2016), 160018. <https://doi.org/10.1038/sdata.2016.18>
- [2] P. Meier-Abt, A.K. Lawrence, L. Selzer, E. Vayena, T. Schwede. The Swiss approach to precision medicine, *Swiss Medical Weekly* (2018). Available from: <https://smw.ch/en/op-eds/post/the-swiss-approach-to-precision-medicine/>
- [3] D. Coman Schmid, K. Cramer, S. Oesterle, B. Rinn, T. Sengstag, and H. Stockinger, SPHN – the BioMedIT network: A secure IT platform for research with sensitive human data, *Studies in Health Technology and Informatics* ?? (2020), ???.

SPHN/PHRT – MedCo in Action: Empowering the Swiss Molecular Tumor Board with Privacy-Preserving and Real- Time Patient Discovery

Jean Louis RAISARO^{a,1}, Juan Ramón TRONCOSO-PASTORIZA^b, Sylvain PRADERVAND^a, Michel CUENDET^a, Mickael MISBACH^b, Joao SA^b, Francesco MARINO^b, Nicolas FREUNDLER^a, Nicolas ROSAT^a, David CAVIN^c, Alexander LEICHTLE^d, Jacques FELLAY^{a,b}, Olivier MICHIELIN^a and Jean-Pierre HUBAUX^b
^a Lausanne University Hospital; ^b EPFL ; ^c Geneva University Hospital; ^d Bern University Hospital

Abstract. MedCo is the first operational system that makes sensitive medical-data available for research in a simple, privacy-conscious and secure way. It enables a consortium of clinical sites to collectively protect their data and to securely share them with investigators, without single points of failure. In this short paper, we report on our ongoing effort for the operational deployment of MedCo within the context of the Swiss Personalized Health Network (SPHN) for the Swiss Molecular Tumor Board.

Keywords. SPHN, personalized health, privacy, security, oncology, data sharing.

1. Introduction

Although within Swiss Personalized Health Network much effort is currently being put into data interoperability and integration, data privacy and security represent important challenges that, if overlooked, can potentially hinder or even block the access and use of this data across the different hospitals. To tackle these issues, researchers from the Swiss Personalized Oncology project and the Data Protection for Personalized Health (DPPH <https://dpph.ch>) project have joined forces. The intention is to work together towards the development and the adoption of cutting-edge privacy-preserving solutions for data analytics that provide strong and long-lasting protection guarantees for patient data.

2. The Swiss Molecular Tumor Board Use Case

As a first step in this direction, the idea is to empower the Swiss Molecular Tumor Board (SMTB), which is meant to be a reference for the entire country for analysing complex cases and identifying the best personalized treatment, with MedCo, the first privacy-

¹ Jean Louis Raisaro, Lausanne University Hospital (CHUV), Chemin des Roches 1, 1010 Lausanne, Switzerland ; E-mail: jean.raisaro@chuv.ch.

preserving tool issued by the DPPH project that is co-developed by EPFL and the Lausanne University Hospital (<https://medco.epfl.ch>). MedCo [1] is a data discovery system that enables a researcher or clinician to securely perform, through a user-friendly graphical interface, exploratory queries (SQL-like queries for simple statistics) on encrypted clinical and molecular data that are horizontally partitioned across different hospitals. By relying on state-of-the-art technologies for information security such as homomorphic encryption, secure multi-party computation and differential privacy, MedCo ensures (1) that patients' individual-level data are protected against illegitimate access by unauthorized people, (2) that no single authority has to be trusted for the security of the data and (3) that the re-identification risk is minimized.

Our goal is to deploy MedCo at all the University Hospitals in order to enable the SMTB to make more informed treatment decisions (Figure 1). MedCo will offer experts of the SMTB the possibility of obtaining, in real time, simple statistics and aggregate information about treatment and outcome for patients with a similar clinical and molecular profile to the ones under evaluation (e.g. overall survival, overall response rate, progression-free survival). At the same time, the use of MedCo will ensure that the privacy of both the patients in the hospitals and the patient presented at the SMTB is preserved. To securely identify similar patients, MedCo will perform an encrypted search over all the oncology patients whose data is stored encrypted in the MedCo databases that are distributed across the five Swiss University Hospitals and then compute the requested aggregate information directly on the encrypted data of the identified patients.

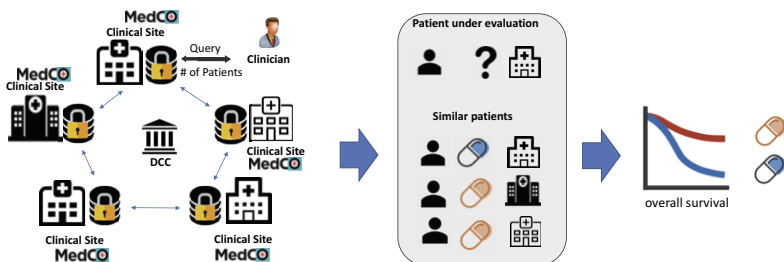


Figure 1. Envisioned use of MedCo in the Swiss Molecular Tumor Board. DCC: Data Coordination Center

3. Conclusion

MedCo will be initially deployed in the University Hospitals of Bern, Geneva and Lausanne for testing purposes. Once the data loading process, the governance and the functionalities of MedCo will be validated by the SPHN Data Coordination Center, the deployment will be extended to the University Hospitals of Basel and Zurich and the other hospitals members of the *Swiss Group for Clinical Cancer Research*, eventually.

References

- [1] J.L. Raisaro, J.R. Troncoso-Pastoriza, M. Misbach, J.S. Sousa, S. Pradervand, E. Missiaglia, O. Michielin, B. Ford, and J.P. Hubaux, MedCo: Enabling Secure and Privacy-Preserving Exploration of Distributed Clinical and Genomic Data, *IEEE/ACM transactions on computational biology and bioinformatics* **16** (2018), 1328-1341.

SPHN/PHRT: Forming a Swiss-Wide Infrastructure for Data-Driven Sepsis Research

Adrian Egli^{a,b}, Manuel Battegay^c, Andrea C. Büchler^c, Peter Bühlmann^d, Thierry Calandra^e, Philippe Eckert^f, Hansjakob Furrer^g, Gilbert Greub^h, Stephan M. Jakobⁱ, Laurent Kaiser^j, Stephen L. Leib^k, Stephan Marsch^l, Nicolai Meinshausen^m, Jean-Luc Paganini^f, Jerome Puginⁿ, Gunnar Rätsch^o, Jacques Schrenzel^p, Reto Schüpbach^q, Martin Siegemund^l, Nicola Zamboni^r, Reinhard Zbinden^s, Annelies Zinkernagel^l, Karsten Borgwardt^u

^a Clinical Bacteriology and Mycology, University Hospital Basel, Basel, Switzerland; ^b Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel; ^c Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, Basel; ^d Department of Mathematics, ETH Zurich, Zurich; ^e Infectious Diseases Service, University Hospital Lausanne, University of Lausanne, Lausanne; ^f Department of Intensive Care Medicine, University Hospital Lausanne, University of Lausanne, Lausanne; ^g Department of Infectious Diseases, Bern University Hospital, University of Bern, Bern; ^h Institute for Medical Microbiology, University Hospital Lausanne, Lausanne; ⁱ Department of Intensive Care Medicine, University Hospital Bern, University of Bern, Bern; ^j Infectious Diseases, University Hospital Geneva, Geneva; ^k Institute for Infectious Diseases, University of Bern, Bern; ^l Intensive Care Medicine, University Hospital Basel, Basel; ^m Department of Mathematics, ETH Zurich, Zurich; ⁿ Intensive Care Medicine, University Hospital Geneva, Geneva; ^o Department of Computer Science, ETH Zurich, Zurich; ^p Division of Bacteriology, University Hospital Geneva, Geneva; ^q Intensive Care Medicine, University Hospital Zurich, Zurich; ^r Institute of Molecular Systems Biology, ETH Zurich, ETH Zurich, Zurich; ^s Institute of Medical Microbiology, University of Zurich, Zurich; ^l Infectious Disease and Hospital Epidemiology, University Hospital Zurich, Zurich; ^u Department of Biosystems Science and Engineering, ETH Zurich, Basel

Abstract. Sepsis is a highly heterogeneous syndrome with variable causes and outcomes. As part of the SPHN/PHRT funding program, we aim to build a highly interoperable, interconnected network for data collection, exchange and analysis of patients on intensive care units in order to predict sepsis onset and mortality earlier. All five University Hospitals, Universities, the Swiss Institute of Bioinformatics and ETH Zurich are involved in this multi-disciplinary project. With two prospective clinical observational studies, we test our infrastructure setup and improve the framework gradually and generate relevant data for research.

Keywords. SPHN, personalized health, sepsis, big data, diagnostics, digital biomarker, -omics biomarkers, machine learning, data exchange, interoperability, interconnected.

1. Introduction

Sepsis is a highly complex, life-threatening syndrome that develops when the bodies' immune response to an infection causes injury to its own tissues and organs. The course and outcome of sepsis is highly heterogeneous and depends on complex host-pathogen interactions, the pathogen, and timing of diagnosis and effective treatment [1, 2]. Patients at risk for sepsis may benefit from personalized diagnostic assessment and treatment strategies. Multiple attempts have failed to develop classical biomarkers from e.g. serum to diagnose sepsis – also due to the heterogeneity of presentation. Digital biomarkers in sepsis may offer a different approach, as the collection of very large datasets allows to integrate time-series data of directly measured physiological parameters [3]. Within an integrative approach, digital biomarkers may be combined with -omics biomarkers (e.g. metabolomics, metagenomic, whole genome sequencing of bacteria, immune phenotyping) and form hybrid biomarkers with higher sensitivity and specificity. The use of novel machine-learning based algorithms to explore such datasets may help to better understand the data and discover patterns within the data linked to particular clinical phenotypes [2]. These algorithms may potentially allow to guide tailored personalized treatment strategies in sepsis. We aim to build a network for data-driven and -omics technology-based research to (i) recognize sepsis at an earlier stage and (ii) predict risk of mortality.

2. Methods

In the driver project funded by the Swiss Personalized Health Network (SPHN) and Personalized Health Related Technologies (PHRT), each University Hospital (Basel, Bern, Geneva, Lausanne and Zurich) and University of Bern and Zurich will provide data from critically ill patients hospitalized in intensive care units (ICUs).

The locally generated data is collected from various hospital information systems (IS) such as clinical and laboratory IS, devices such as ventilators and dialysis machines, ICU monitors, and treatment-related data, e.g. drug doses and transferred and stored in local clinical data warehouses (CDWH). In order to follow Sepsis 3.0 criteria [4], we regularly tag patient with suspected sepsis and collect information in the clinical IS. This will allow to compute Sepsis-3 criteria and compare data patterns to this international accepted definition of sepsis. The data is structured and interoperable between centres following SNOMED CT and LOINC ontologies. Data is then transferred in a resource description framework (RDF) format after encoding and encryption via the BioMedIT nodes to the ETH domain for further analysis. A shared data model will be used for machine-learning based analysis. The legal and governance framework is established using a Data Transfer and Use Agreement (DTUA) and Consortium Agreement (CA) [Figure 1].

The clinical outcomes of our study include: (i) prediction of sepsis with new markers in comparison to Sepsis-3 criteria and (ii) prediction of in-hospital mortality. There are two prospective observational trials conducted with these patients.

Study A includes every patient admitted to an ICU. Following specific inclusion/exclusion criteria, we include sepsis (cases) and non-sepsis (controls) upon entry of the ICU. Of all patients with an ICU admission also available pre/post-ICU data will be integrated into the data analysis. We will compare data patterns between cases and controls in regards of the previously defined outcomes.

Study B includes a subgroup of the patients from Study A. In all five hospitals a total of 300 patients with severe community acquired pneumonia (cases) and 100 patients with a severe systemic inflammation (controls) will be included. We will collect additional patient material such as serum, whole blood for DNA and RNA, respiratory and stool material, immune cells and bacterial isolates. These samples will be in-depth characterized using metabolomics, metagenomics, whole genome sequencing and immune phenotyping technologies. Digital and -omics data will then be merged and analysed to form hybrid biomarkers. Again, case and control patients will be compared in regards of the previously defined outcomes.

3. Results

For the digital biomarker project (Study A) we have obtained ethical approval to collect data from 15,000 ICU patients from participating centres over the next few years. In addition, a data transfer and use agreement as well as a consortium agreement have been generated. The legal process to allow data sharing was complex and involved a multiple round assessment and discussion of the legal framework with various stakeholders.

A list of clinical, physiological and laboratory variables, covering more than 500 attributes of interest was generated and linked to ontologies such as time points, application route and dosage of specific drugs, etc. Collection of data has started at local levels with a total of already several hundred patients from the participating ICUs. Within the next 6 months, we expect data from 3,000 patients to be collected.

Data transfer and exchange protocols are currently tested with all CDWH, BioMedIT, and ETH domain teams and continuously improved.

The second -omics biomarker project (Study B) has not yet started. The ethical proposal has been submitted and decision is pending. Study preparation are progressing at all centers. In regards of the sample management different biobank information management software (BIMS) systems have been established in the various centers.

4. Conclusions

Ethical and legal frameworks are critical in a data-driven large consortium and this bottleneck should not be underestimated. The legal discussions show a different momentum, content and motivation compared to research questions. This should be anticipated in such a project. A lead legal institution may be a good concept to provide legal advice, besides the availability of the DTUA and CA. The legal situation of large healthcare project is complex and requires a professional support. There is a clear benefit of regular meetings with all stakeholders for in-depth discussions and finding solutions for challenges. Once fully functional, the current framework will allow to collect and store data for research purposes across all Swiss University Hospitals ICUs – beyond sepsis.

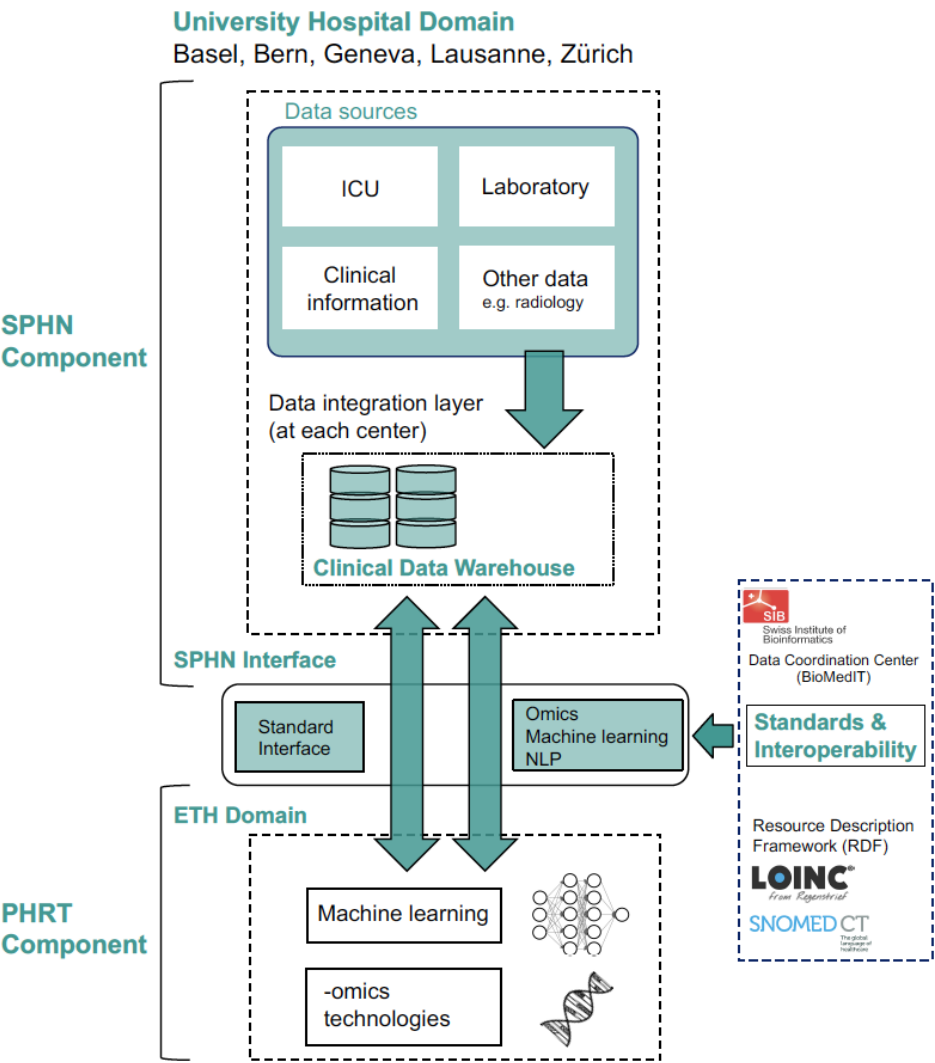


Figure 1. SPHN/PHRT framework for the Sepsis driver project. All centers locally collect data from different primary sources in clinical data warehouses. The is high interoperable and standardized following ontologies. The exchanged follows secured routes of BioMedIT to the ETH Domain for quality control and analysis. [1]

Acknowledgements

This project was supported by the Swiss Personalized Health Network (SPHN) initiative and the Personalized Health and Related Technologies (PHRT) initiative (Grant number: 2017DRI20 [PHRT110]).

References

- [1] A. Kumar, D. Roberts, K.E. Wood, B.A. Light, *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock, *Crit Care Med* **34**, (2006), 1589-96.
- [2] C.W. Seymour, J.N. Kennedy, S. Wang, C.H. Chang, *et al.* Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis, *JAMA* **321** (2019), 2003-2017.
- [3] M. Komorowski, L.A. Celi, O. Badawi, *et al.* The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care, *Nat Med* **24** (2018), 1716-1720.
- [4] M. Singer, C.S. Deutschman, C.W. Seymour, M. Shankar-Hari, *et al.* The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3), *JAMA* **315** (2016), 801-810.

SPHN – The Swiss Aging Citizen Reference (SACR)

Ayoung JEONG^{a,b}, Murielle BOCHUD^c, Philippe CATTIN^b, Manolis DERMITZAKIS^d, Bogdan DRAGANSKI^e, Andreas PAPASSOTIROPOULOS^f, Martine PREISIG^c, Bram STIELTJES^c, Peter VOLLENWEIDER^c, and Nicole PROBST-HENSCH^{A,b,1}

^aSwiss Tropical and Public Health Institute; ^bUniversity of Basel;

^cUniversity Hospital Lausanne; ^dUniversity of Geneva; ^eUniversity Hospital Basel;

^fTransfaculty Research Platform Molecular and Cognitive Neurosciences

Abstract. In Switzerland by 2045, we expect 2.7 Mio citizens aged 65+ of whom 1.0 Mio. aged 80+. A priority and focus of personalized health research is therefore aging biology to extend healthy life expectancy. Novel molecular and imaging features will emerge as candidate targets for risk prediction and screening of chronic diseases. It is of utmost importance to test the clinical and public health utility of candidate biomarkers evolving from this research in citizen reference cohorts. We will build a Swiss Aging Citizen Reference (SACR), a testable and scalable reference cohort offering interoperable, searchable, and accessible data. 1000 participants from existing Swiss citizen cohorts will be combined and analyzed for DNA methylation and MRI brain imaging. SACR will serve as a testbed for clinical and public health utility of candidate biomarkers. As for a proof-of-concept study, we will conduct an agnostic search for structural and functional brain features associated with epigenetic aging acceleration to examine the potential of epigenetic age acceleration as the intermediate aging biomarker and to better understand the aging mechanism in brain.

Keywords. SPHN, personalized health, citizen reference, aging, DNA methylation, brain imaging

1. Introduction

The number of older citizens in Switzerland is growing. In 2045 pensioners aged 65+ are expected to account for over 2.7 million inhabitants, with over 1 million aged 80+. It is of strategic importance for the country from both, economic and public health perspectives, to maximize the healthy life expectancy of future elderly people.

Molecular and imaging biomarkers evolving from personalized health research depend on a healthy reference for utility testing. Citizen cohorts with prospectively sampled biomaterial are essential for investigating the reference distribution of exposome, risk prediction and screening biomarkers and their independent value in predicting morbidities. Among many indices and markers of biological aging, DNA methylation in blood has been demonstrated as a robust aging biomarker, with potential to interrogate molecular mechanism behind aging. DNA methylation-derived epigenetic

¹ Corresponding Author, E-mail: nicole.probst@swisstph.ch.

age acceleration was cross-sectionally associated with poorer cognitive performance (fluid intelligence), lower grip strength, poorer lung function, and with frailty.

Cognitive decline and neurological disorders are becoming one of the most significant socio-economic burdens of our aging society. One of the fundamental questions in human neuroscience is how molecular and cellular processes in the living brain drive cognitive changes over the lifespan. In recent decades the study of brain structure and function in health and disease has been greatly aided by the advent of magnetic resonance imaging (MRI) and started revealing how acute focal lesions and chronic neurodegeneration can lead to impaired cognition and disease states.

In this SPHN driver project we will build a Swiss Aging Citizen Reference (SACR) as a public health pillar for personalized health in Switzerland, where we implement brain imaging biobank into existing Swiss citizen cohorts to test the clinical and public health utility of candidate biomarkers including epigenetic age acceleration.

2. Methods

We build a scalable SACR infrastructure to integrate existing Swiss citizen cohorts, SAPALDIA, SKIPOGH, and CoLaus|PsyCoLaus, offering interoperable, searchable, and accessible data for researchers. The SACR population will continue to grow by integrating additional citizen cohorts in the future.

DNA methylome and MRI brain imaging data will be obtained for 1000 SACR citizens from the three cohorts. DNA methylome is already available in SAPALDIA and will be obtained for SKIPOGH and CoLaus|PsyCoLaus. MRI brain imaging has been conducted in CoLaus|PsyCoLaus and will be complemented for SAPALDIA and SKIPOGH. As for a proof-of-concept study, structural and functional features derived from quantitative MRI indicative for brain's myelin, iron and free tissue water content will be agnostically searched for association with epigenetic age acceleration.

3. Expected results

From the proof-of-concept study, we will identify brain microstructure features associated with epigenetic age acceleration, which is considered as an early biological aging marker, therefore we will provide evidence towards the validity of epigenetic age acceleration as intermediate aging biomarker. Further analyses on the brain imaging features towards aging phenotypes, e.g. cognitive function, will contribute to better understanding of brain aging, and potentially to its prediction at an early stage.

4. Public health impacts

As a result of this SPHN driver project, we will have a testable and scalable reference “SACR”. This will constitute the public health pillar for personalized health in Switzerland, offering interoperable, searchable, and accessible data for personalized health researchers. Archived biomaterial and data from SACR participants will be accessible for testing the clinical and public health utility of candidate biomarkers and exposome features in predicting biological age and age-related (multi-)morbidity.

SPHN – The BioMedIT Network: A Secure IT Platform for Research with Sensitive Human Data

Diana COMAN SCHMID^{c,a}, Katrin CRAMERI^{a,1}, Sabine OESTERLE^a, Bernd RINN^{c,a}, Thierry SENGSTAG^{b,a} and Heinz STOCKINGER^a
on behalf of the BioMedIT network team: Leila T. ALEXANDER^a, Jonathan BARDA^a, Christian BOLLIGER^{c,a}, Urban BORSTNIK^c, Gerhard BRÄUNLICH^{c,a}, Olivier BYRDE^c, Jérôme DAUVILLIER^a, Robin ENGLER^a, Pablo ESCOBAR LOPEZ^{b,a}, Volker FLEGEL^a, Martin FOX^a, Sofia GEORGAKOPOULOU^{b,a}, Jani HEIKKINEN^{b,a}, Martin JACQUOT^{b,a}, Nicolas KOWENSKI^c, Guillermo LOSILLA^{b,a}, Sergio MAFFIOLETTI^{c,a}, Jorge MOLINA^a, Diego MORENO^c, Allen NEESER^c, Michal OKONIEWSKI^{c,a}, Warren PAULUS^a, Kevin SAYERS^a, Torsten SCHWEDE^{b,a}, Jaroslaw SURKONT^{b,a}, Richard WARTENBURGER^c, and Thomas WÜST^{c,a} (in alphabetical order)

^a*SIB Swiss Institute of Bioinformatics, Switzerland,*

^b*University of Basel, Switzerland,*

^c*ETH Zurich, Switzerland*

Abstract. The BioMedIT project is funded by the Swiss government as an integral part of the Swiss Personalized Health Network (SPHN), aiming to provide researchers with access to a secure, powerful and versatile IT infrastructure for doing data-driven research on sensitive biomedical data while ensuring data privacy protection. The BioMedIT network gives researchers the ability to securely transfer, store, manage and process sensitive research data. The underlying BioMedIT nodes provide compute and storage capacity that can be used locally or through a federated environment. The network operates under a common Information Security Policy using state-of-the-art security techniques. It utilizes cloud computing, virtualization, compute accelerators (GPUs), big data storage as well as federation technologies to lower computational boundaries for researchers and to guarantee that sensitive data can be processed in a secure and lawful way. Building on existing expertise and research infrastructure at the partnering Swiss institutions, the BioMedIT network establishes a competitive Swiss private-cloud – a secure national infrastructure resource that can be used by researchers of Swiss universities, hospitals and other research institutions.

Keywords. Personalized health, SPHN, research infrastructure, scientific IT services, federated computation, sensitive data, confidential data, data privacy, data security, health-related data, interoperability, private-cloud, service virtualization

¹ Corresponding Author: Katrin Cramer, Personalized Health Informatics Group, SIB Swiss Institute of Bioinformatics, Elisabethenstrasse 43, 4051 Basel, Switzerland; E-mail: katrin.cramer@sib.swiss

1. Introduction

The advent of digital transformation in health care has produced an exponential increase in the amount of information available for each patient. The use of this information in data-driven biomedical research has the potential to drive important changes in medicine [1]. In order to be able to leverage the potential of health-related data through biomedical research, data science and related research fields, data needs to be interoperable and available to researchers in various disciplines. In addition, strong capabilities in clinical bioinformatics and computational service infrastructure are required in order to enable the integration and interpretation of large and rich data sets. Furthermore, big data analyses and machine learning require vast amounts of data and accordingly high-performance IT infrastructures for computing and storage.

Given the sensitive nature of health-related information, data-driven personalized health research requires special IT infrastructures and services, blending the concepts of (i) security and compliance to protect the confidentiality of the data and the privacy of the research study participants; (ii) scalability and performance to be able to adapt to changing needs of users; and (iii) flexibility and ease of use to foster cutting-edge biomedical research. Security measures for Information and Communications Technology (ICT) systems are necessary to protect confidential information from unauthorized use, modification, loss or release². A few years ago, these requirements were not a major concern for academic computing facilities in Switzerland because those were predominantly tailored towards the handling of (insensitive) basic research data.

Another important requirement concerning the architecture of IT infrastructures for researchers working in multidisciplinary networks on big data sets in the biomedical field, is the possibility for controlled access to data sets across research teams from different institutions, also cross borders. In addition, in the context of nation-wide collaborative research projects, technical interoperability between different IT infrastructures should be granted in order to enable reproducibility of data analysis workflows executed at distributed locations [3].

To address the needs listed above, the BioMedIT project was funded by the Swiss federal government for the period of 2017-2020 within the framework of the Swiss Personalized Health Network Initiative (SPHN) [4] and in close collaboration with the strategic focus area Personalized Health and Related Technologies (PHRT) of the Swiss Federal Institutes of Technology (ETH) domain [5]. The aim of the project is to provide all researchers in Switzerland with access to a service infrastructure for collaborative analysis of confidential data without compromising data privacy. The intention is to create and maintain a national infrastructure resource that can jointly be used by all Swiss universities, research institutions, hospitals and other interested partners.

2. The BioMedIT Network: a National Secure Infrastructure Resource for Switzerland

The BioMedIT network builds on three scientific IT competence centers – the BioMedIT nodes – in different geographical locations: one in Basel (sciCORE, operated by the University of Basel), one in Lausanne (Core-IT, operated by the SIB Swiss Institute of

² Key elements of an effective ICT security system include (i) Monitoring and controlling access to confidential information; (ii) Safe transmission of data, and (iii) Secure storage and disposal of data [2].

Bioinformatics) and one in Zurich (SIS, operated by ETH Zurich). Over the past two years, all three nodes established secure compute and data platforms specialized on handling confidential research data that is subject to legal and ethical constraints. The computational service infrastructure is available to Swiss universities, research institutes, hospitals, service providers and other interested partners (*organizational clients*, see Section 2.4) that are not part of the BioMedIT network, to securely store, manage and process confidential research data. In addition, and specifically within the framework of the two Swiss national initiatives SPHN and PHRT, the BioMedIT network can be used for mono- and multi-site, individual and collaborative research projects (*project clients*, see Section 2.3). In fact, several innovative biomedical research projects [6] are already using a first version of the BioMedIT infrastructure.

2.1. Security Aspects of the BioMedIT Network

A common Information Security Policy [7] that applies to the entire BioMedIT network, defines the necessary organizational and technical measures to allow researchers to process sensitive data in a secure way³. This policy controls the way the confidentiality, integrity, and availability of information is handled, preventing misuse and malicious damage. In conventional high performance or high throughput systems available in scientific environments, the physical hardware and the respective software services are shared between all users. For data privacy and security reasons, BioMedIT takes a different approach: researchers involved in one particular project work completely isolated from any other project, meaning that all the data, software and resources belong exclusively to that project and there is no possibility to access it in any other way or from any other path. To this end, virtualization technologies are applied to establish private-cloud environments in which each scientific project is completely isolated from other projects with respect to data and compute resources. Access of authorized users to any project space requires two-factor authentication. Furthermore, users that are authorized to use the secure BioMedIT network can only access the infrastructure from within trusted IT environments (either from within Swiss university or university hospital networks or via VPN). Access to the Internet from the BioMedIT nodes is strictly controlled, limited to trusted and explicitly white-listed web resources.

The three BioMedIT nodes share a common security architecture to transfer, store, manage, analyze and share biomedical data while following the latest technical and legal standards required by Swiss legislation⁴ as well as internal regulations of the associated institutions. For project-related data transfer, an end-to-end encryption process from the data source through the BioMedIT network to the project space is set up, based on public-key cryptography. Special consideration is given to key management which is provided as a central service to users of the network.

Furthermore, the network follows the SPHN Ethical Framework for Responsible Data Processing in Personalized Health Research [11]. Importantly, before using the secure BioMedIT network, users are requested to follow a data security awareness training. BioMedIT offers the “Data Privacy and IT Security Training” as on-line

³ In general, sensitive data (as it is defined according to Swiss and international data protection legislation) is classified as “confidential” according to the SPHN Information Security Policy.

⁴ In particular, the Federal Act on Data Protection [8], the Human Research Act [9], the EU General Data Protection Regulation [10]

training [12] or class-room course hosted regularly in different cities in Switzerland. Finally, users must pass a mandatory on-line exam.

2.2. Technical Features of BioMedIT's ICT environment

The BioMedIT nodes provide a broad range of secure services to store, manage, process and share biomedical data. For example, multi-CPU (central processing unit) and multi-GPU (graphics processing unit) compute nodes are connected by fast internal networks to large and fast parallel file systems and state-of-the-art data management services. To date, the BioMedIT infrastructure features data storage capacity approaching in total 5 PB, support for data encryption, secure backup, private-cloud and high performance computing (HPC) leveraging more than 3000 CPU cores and 324 GPUs. State-of-the-art software for data science and specialized tools for data management are provisioned alongside with long-standing expertise in scientific IT support for research data management, bioinformatics, HPC and computational analysis.

2.3. Lowering Computational Barriers for Researchers (Project Clients)

The BioMedIT network provides a flexible ICT environment, and individual project spaces can be configured based on the researcher's computational and storage needs. This includes storage and compute capacity, up-to-date data analysis and modeling software packages, as well as backup and archiving solutions. The nodes can provide secure work environments for basic exploratory data analysis but can also scale up to cover high performance and high throughput needs such as large-scale machine learning.

The BioMedIT portal, a central service of the network operated by the SPHN Data Coordination Center of SIB's Personalized Health Informatics Group, provides a single access point to the BioMedIT nodes and associated resources, simplifying access control for the researchers. Depending on the use-case, the BioMedIT network can be accessed by command-line or by web-based remote-desktop technology.

Standardized ways for secure data transfer from partner institutions, such as the five Swiss University Hospitals, the PHRT data centers and other data providers are currently being established and will be available to all projects relying on data from any of these providers.

The BioMedIT network aims to enable interoperable workflow execution, providing a way for researchers to work seamlessly across the nodes. Containerization of data workflows using Open Container Initiative (OCI) standards [13] are playing an important role for this and will at the same time improve reproducibility of results obtained by these workflows. The ultimate goal of BioMedIT is to provide a data-aware federated analysis platform where researchers can work with distributed data. This approach is especially valuable for confidential data or large data sets, which cannot easily be shared.

2.4. Outsourcing Sensitive Research Data Operation to BioMedIT (Institutional Clients)

While the focus of the BioMedIT network is to support national research projects in the SPHN and PHRT programs, the BioMedIT nodes are creating a unique data and computing environment that will be useful and potentially required for research activities outside of these programs. Clinical institutions scaling up their data processing in preparation of clinical research projects or clinical decision support strategies,

consortium projects with public and private partners, or novel ways to run clinical trials, which create large amounts of trial data and require good clinical practice (GCP) compliant data management, are examples for it. The BioMedIT nodes are offering a safe environment for these types of activities and are working on enhancing their platforms for future needs. By linking to the BioMedIT network, Swiss research institutions and hospitals can hence benefit from a trusted research IT infrastructure without the need to (re)develop costly *in-house* infrastructures and know-how.

3. Conclusion

BioMedIT provides data resources and a network of secure data nodes for doing data-driven research on sensitive (confidential) biomedical data. It lowers computational barriers by provisioning an efficient solution for cross-institutional analysis of sensitive data, and is therefore unique in Switzerland. The technical solution is comparable to related activities within ELIXIR in various European countries [14]. However, BioMedIT operates within a national network of health-care providers, infrastructure and research support groups, research projects as well as citizens, and therefore has a great potential to become a platform of reference for health data exchange and analysis. A first version of the BioMedIT network is established and already used by SPHN and PHRT projects and pilot collaborations with Swiss university hospitals. These projects help shaping new versions of the BioMedIT infrastructure to make it even more useful, efficient and secure for current and future biomedical research applications.

References

- [1] C. Lovis, Unlocking the Power of Artificial Intelligence and Big Data in Medicine, *J Med Internet Res.* **21** (2019), e16607.
- [2] Australian Government, A Guide for Data Integration Projects Involving Commonwealth Data for Statistical and Research Purposes <https://statistical-data-integration.govspace.gov.au/topics/secure-data-management/information-and-communication-technology-security>, accessed 17 January 2020
- [3] Enhancing Research through IT Expertise <https://www.enhancer.ch/pipeline>, accessed 17 January 2020
- [4] The Swiss Personalized Health Network Initiative www.sphn.ch, accessed 17 January 2020
- [5] Personalized Health and Related Technologies, <https://www.sfa-phrt.ch>, accessed 17 January 2020
- [6] SPHN Projects, Bottom-up funding: Driver Projects and Infrastructure Development Projects <https://sphn.ch/network/projects>, accessed 17 January 2020
- [7] SPHN Information Security Policy, Version 1.0 of 23 August 2018, https://sphn.ch/wp-content/uploads/2020/01/sphn_information_security_policy_v1.pdf
- [8] Federal Act on Data Protection of 19 June 1992 (Status as of 1 March 2019), SR 235.1 <https://www.admin.ch/opc/en/classified-compilation/19920153/index.html>
- [9] Federal Act on Research involving Human Beings of 30 September 2011 (Status as of 1 January 2014), SR 810.30 <https://www.admin.ch/opc/en/classified-compilation/20061313/index.html>
- [10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>
- [11] Ethical Framework for Responsible Data Processing in Personalized Health Research, Version 2, of 7 May 2018 https://sphn.ch/wp-content/uploads/2019/11/Ethical_Framework_20180507_SPHN.pdf
- [12] <https://edu.sib.swiss/course/view.php?id=424>, accessed 17 January 2020
- [13] Open Container Initiative <https://www.opencontainers.org>, accessed 15 January 2020
- [14] ELIXIR <https://elixir-europe.org/>, accessed 17 January 2020

Section 8

Posters

This page intentionally left blank

A Digital Service Logistics Information System for Emergency Department Care Coordination - Professionals' Experiences

Laura-Maria PELTONEN^{a,1}, Armi KULOHEIMO^a, Kristiina JUNTILA^a, Sanna SALANTERÄ^a

^a*Department of Nursing Science, University of Turku, Turku, Finland*

Abstract. The aim was to describe nurses' and physicians' perceptions of a digital service logistics information system from an operative management perspective in emergency departments (EDs). A total of 24 professionals were interviewed. Based on the results the information systems support operative management of EDs but the professionals desire more detailed information about patients and staff to support situational awareness in the operative management of these units.

Keywords. Emergency Department, Information System, Care Coordination

1. Introduction

Although information management in health care today is widely digitalized and electronic health records have received attention, less is known about information systems developed for operational management. Timely access to important information is essential when managing the day-to-day operations of emergency departments (EDs) in hospitals as decisions typically must be made without delay [1].

This study focuses on a service logistics information system (Columna Clinical Care Logistics®) that displays information about patients, staff and care processes for the professionals. It is available in hospitals on screens, personal computers and mobile devices. The system aims to support easy access to important information for multiprofessional needs. The aim of this study was to describe nurses' and physicians' perceptions about this information system from an operative manager's perspective in an ED.

2. Methods

The study had a cross-sectional design. Data were collected in three EDs, including two emergency departments and one emergency inpatient ward in two central hospitals in Finland. A total of 24 professionals, including 7 physicians and 17 nurses, who were in charge of coordinating care, participated in the study. Data were collected through interviews with two open ended questions: 1) What do you think is good about the

¹ Corresponding Author, E-mail: laura-maria.peltonen@utu.fi.

service logistics system? and 2) what could be improved regarding the service logistics system? Data were collected in the autumn of 2018 and spring of 2018. Data were analysed using thematic content analysis [2].

3. Results

The mean age of the respondents was 42.4 years (SD 10.2) and their mean work experience was 17.2 years (SD 11.7). Three respondents worked only at office hours, three worked only beyond office hours and eighteen worked both during office hours and beyond. Issues regarding the information system were divided into three main themes.

- Support for care coordination: The system provides multidisciplinary task management support and a quick overview of patients, staff and processes, but the complexity of care needs and staff skills was not shown in sufficient detail.
- System implementation and usage: Personal traits and the information management culture influenced system usage, and professionals used the system to a different extent when measured at one year after implementation.
- Information exchange: A lack of information exchange with other systems limited the possible benefits in communication between care providers.

4. Discussion

The digital service logistics information systems support operative management of EDs. However, professionals desire more detailed information about patients and staff for situational awareness. Artificial intelligence could be used to support this task. Further, it takes a long time to learn how to use such an information system as it changes the professionals' work. Hence, the need of a long follow-up period needs to be considered when measuring the impact of such systems to ensure that all users use the system in the way expected. In conclusion, digital service logistics information systems can support professionals' situational awareness in EDs, but they should be developed based on evidence of information needs. Reliable use requires that all users use the system to its full potential.

References

- [1] H. Lundgrén-Laine, E. Kontio, J. Pertilä, H. Korvenranta, J. Forsström, S. Salanterä, Managing daily intensive care activities: an observational study concerning ad hoc decision making of charge nurses and intensivists. *Critical Care* 15 (2011), 1–10.
- [2] U.H. Graneheim, B.A.N. Lundman, Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness, *Nurse Educ Today* 24 (2004), 105–12.

A Matlab Tool for Organizing and Analyzing NHANES Data

Simon Lebech CICHOSZ¹, Morten Hasselstrøm JENSEN^{1,2},
Thomas Kronborg LARSEN¹, Ole HEJLESEN¹

¹*Department of Health Science and Technology, Aalborg University, Denmark*

²*Steno Diabetes Center North Denmark, Aalborg University Hospital, Denmark*

Abstract: Automation of organizing and analyzing NHANES data can provide easier access to data and potentially reducing risk of introducing bias. This study investigates the potential for developing a software for this purpose. MATLAB R2016b was used for transforming and analyzing data from the NHANES. The software was tested successful by analyzing the association between smoking and glucose metabolism in the general population.

1. Introduction

The National Health and Nutrition Examination Survey (NHANES) (1) is a large, stratified, multistage survey conducted by the Centers for Disease Control (CDC) which collects health data on the US population. Data are made publicly available in two-year cycles and the data are categorized into six main areas; dietary, demographics, examination, laboratory, questionnaire, and limited access data.

However, there is an absence of reproducible software and methods that transforms the data from its published online format into an analytic dataset. Even though the NHANES data are publicly available, actually analyzing the data requires a non-trivial amount of background information, data processing and linking. This preprocessing is prone to introducing bias into the data and requires considerable time and resources. We sought to investigate the potential for developing a software for simplifying this process and as an example applying the tool for analyzing the association between smoking and glucose metabolism in the general population.

2. Methods

We developed in MATLAB R2016b a two-level tool. Based on the desired NHANES year-cycles and data variables of interest the first level of the tool will download the relevant data from the public repositories and organize the files locally. The data files are published as a .xpt (SAS XPORT) format, therefore the files are subsequently imported and converted to tables in MATLAB which are then linked based on year-cycle and type of variable. Based on the merged data the second level of the tool is able to analyze data (using an chosen statistical method) and produce tables with patients' characteristics and effect estimates of independent variables in an unadjusted or adjusted analysis. In our study of smoking and glucose metabolism, several years of NHANES (2005-2014) and data types were utilized. From demographics, gender and age data were

collected; from laboratory, fasting glucose and oral glucose tolerance test (OGTT) results were collected; from examination, blood pressure and body mass index (BMI) were collected; and from questionnaire, data on smoking and alcohol habits were collected. The effect of current smoking was assessed in an unadjusted and an adjusted analysis, taking the mentioned covariates into account (gender, age, alcohol usage, BMI).

3. Results

The MATLAB tool was successful in organizing and analyzing data from several year-cycles of NHANES. A total of 12460 participants was identified as currently smoking (n=2575), ex-smokers (n=3032), and never smoking (n=6853). Currently smoking was in the unadjusted analysis associated with higher fasting glucose (effect 3.46 mg/dL, SE:0.72, $p<0.001$). In the adjusted analysis smoking was associated only with lower OGTT (effect -3.94 mg/dL, SE: 1.44, $p=0.01$)

4. Discussion and conclusion

Automation of organizing and analyzing NHANES data can provide easier access to data and potentially reducing risk of introducing bias. The tool was successfully applied on data from several year-cycles. The test results showed a surprisingly negative association between smoking and OGTT. This association needs to be investigated in more detail.

5. References

- [1] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2005-2014

A Path to Inclusion: Design and Prototype of Transgender Identity in an Electronic Health Record

Melanie CASSARINO^{1a}, Eunice CORREA^a, Sebastián MINOLETTI^a, Ignacio JAUREGUI^a, María Florencia GRANDE RATTI^a, Carlos OTERO^a, Daniel LUNA^a

^a*Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina*

Abstract. With the purpose of reducing discrimination on gender, the Department of Health Informatics of the Hospital Italiano de Buenos Aires (HIBA) developed changes in its Patient Master Index to register chosen name and self-perceived gender identity. To improve the visualization of information related to the identity of each person to provide comprehensive quality care, it was decided to modify the impact of this information on the EHR. For this, prototypes were developed with changes in the patient identification module and then user-tests were carried out.

Keywords. Transgender, Electronic Health Records, Identification, Gender.

1. Introduction

Gender identity is the internal and individual experience of gender as each person feels it, which may or may not correspond to the sex assigned at the time of birth. This may involve the change of the appearance or body function through pharmacological agents, surgery or other, when it is freely chosen. It also includes other gender expressions, such as dress, speech and manners.

Since 2018, the **self-perceived gender** of the patient and their **chosen name** are included in the registration system [2]. In order to improve the visualization of information related to the identity of each patient to provide comprehensive quality care, it was decided to incorporate this information on the EHR. For this, prototypes were developed with changes in the patient identification module and then user-tests were carried out. The aim of this study is to describe the user-centered design process, used to visualize the identification of transgender patients within the Electronic Health Record.

2. Methods

Inquiries and usability tests were carried out with HIBA health professionals. Sample selection was for convenience. Users had to present the following inclusion criteria: be a health professional (regardless of training level), have used the EHR for at least 6 months in in- and outpatient settings, without knowledge of the Transgender project.

¹ Melanie Cassarino, E-mail: melanie.cassarino@hospitalitaliano.org.ar

2.1 Design

Stage 1: Situational diagnostic: Includes reviewing the literature using PUBMED. The articles were manually reviewed and a selection was made of those that mentioned the identification of the different identities in the EHR. Then, the current legislation in our country was evaluated to include the regulatory framework.

Stage 2: Design and evaluation: Prototype Design: From the bibliography found, the Gender Identity Law [1] and the iconography, different prototypes were designed to cover the different use cases. They consisted of a capture of the summary module of the Electronic Health Record. On the sidebar, where the affiliate data is found, the icon was linked with the chosen name, followed by the initials of the legal name. Likewise, the last name, the chosen name, the legal name, the gender, a hyperlink to the gender law and the biological sex were included in the "More data" menu.

Stage 3: User tests: A test round was carried out with five different users. Interfaces of the EHR with the iconographic incorporation were evaluated. A clinical case about a transgender patient with a chosen name, but without a change of birth ID, was conducted. Users were shown a mock up with the visualization of the case data. They were asked to interpret the patient from data, describe what they saw, decide the gender of the patient. Each professional was asked about whether he detected any change in the display of the patient's identity information, what the selected icon meant and what he thought the letters in brackets were referring to. This test was carried out in the workspace of each professional. After each test, reports were made with the findings per user.

3. Results

Prototype testing: All focused their answers on the patient's **clinical** data, that is, about the diagnosis, the reason for consultation and medical history. The majority of the professionals who participated in the test interpreted without hesitation that the patient in the case was "male". Only two users doubted the gender because it presented interconsultation with gynecology. This was expressed in phrases such as "*I doubt because in records it says that it consults for infection of the cervix, is he transgender?*" and "*I think it is male because his name is Mariano but here in records they put it that it has a cervix and also they made him an interconsultation with gynecology. Did they make a mistake?*"

All professionals agreed that they usually infer the patient's sex through their name. The majority could recognize, after careful observation, that the icon represented in the prototype corresponded to the "transgender" identity. In turn, all users agreed that the letters in parentheses were new information in the EHR, but could not infer their meaning. These users expressed "*I don't know, but now that I see that it is transgender I think that maybe is related to its old name*". In conclusion, we believe that it is essential to accompany the adapting process to the information system with instances of awareness, education and training of both administrative and health staff. In this sense, it is vital that all employees of the institution understand the importance of respecting each person's perception of themselves and how they want to be identified. Respecting and enforcing each other's gender identity allows us to live in a society with less discrimination and more understanding of others.

A Recommender System Based on Cohorts' Similarity

João Rafael ALMEIDA^{a,b}, Eriksson MONTEIRO^c, Luís Bastião SILVA^c, Alejandro PAZOS^b, and José Luís OLIVEIRA^a

^aDETI/IEETA, University of Aveiro, Portugal

^bDepartment of Computation, University of A Coruña, Spain

^cBMD Software, Portugal

Abstract. Aiming to better understand the genetic and environmental associations of Alzheimer's disease, many clinical trials and scientific studies have been conducted. However, these studies are often based on a small number of participants. To address this limitation, there is an increasing demand of multi-cohorts studies, which can provide higher statistical power and clinical evidence. However, this data integration implies dealing with the diversity of cohorts structures and the wide variability of concepts. Moreover, discovering similar cohorts to extend a running study is typically a demanding task. In this paper, we present a recommendation system to allow finding similar cohorts based on profile interests. The method uses collaborative filtering mixed with context-based retrieval techniques to find relevant cohorts on scientific literature about Alzheimer's diseases. The method was validated in a set of 62 cohorts.

Keywords. Alzheimer cohorts, Cohort catalogue, Recommendation systems

1. Introduction

There is a need for discovering new biomarkers to diagnose Alzheimer's disease in the predementia stage to predict the rate of decline [1]. Therefore, several institutions increased the efforts in collecting Alzheimer's patients' data, as well as to build a harmonised structure to store these cohorts.

In the EMIF project¹, we have developed a platform to collect and publish metadata about EHR databases and disease specific cohorts [2]. The EMIF Catalogue includes several communities, one of them dedicated to the study of Alzheimer's disease. In this centralised platform, data sets are classified through a set of general characteristics, which allow the identification of similar patterns across the collection. However, the manual analysis of these similarities is time consuming and hard to perform.

In this paper, we propose a methodology that, based on previous knowledge gathered from user's preferences and user's data sets, is able to suggest similar cohorts and related scientific publications. The method uses collaborative filtering and context-based retrieval techniques, taking as inputs the cohorts' metadata in a predefined ontology, its concepts and relationships.

¹<http://www.emif.eu>

2. Methods and Materials

The EMIF Catalogue is a web platform for biomedical data sharing, where data custodians can publish information about their data sets [2]. This information is mainly statistical information and aggregated databases' metadata. In the platform, researchers can search for databases of interest and request access to the raw data, data subsets, or ask to perform certain research questions.

Most of the research studies do not follow a standard structure or format, because they are built with a specific purpose, with inclusion/exclusion criteria. To allow the reproducibility of research questions in different cohorts, our strategy was the adoption of a taxonomy implemented using an ontology with the recorded variables and also the values (columns and rows). This method allows comparing multiple cohorts, i.e., all the variables will be mapped into ontology classes, and this ontology supports relationships between concepts, following a hierarchical tree with root entities that represent core categories that are being followed (i.e. Patient Demographic Data, Neuropsychiatry, Laboratory Results, etc.). Moreover, it is also possible to create constraints in the cohort scope, and it is flexible enough to extend or create new variables. The ontology management is maintained by community using a common RDF (Resource Description Framework) specification.

3. Results and Conclusions

We proposed a recommendation system that combines two techniques. The collaborative filtering can detect similar users and provide the recommendation based on the analysis when the cohorts structure of the users interest are too disperse. On the other hand, the context-based retrieval can predict suggestions when the user is more singular, by relying only in the cohorts' concepts. We applied evaluation metrics to each individual approach and then combining both strategies. The methodology was integrated in the EMIF Catalogue platform, using the statistical data collected from the Alzheimer's community. This group gathered metadata information from 62 cohorts, representing in total 661 concepts, which have also been indexed through our system. This community has more that 500 registered users, allowing a first test-bed to apply our methodology.

Acknowledgments

This work was also funded by the project POCI-01-0145-FEDER-016385 (NETDIAMOND) co-funded by Centro 2020 program, Portugal 2020, European Union, through the National Science Foundation. JRA is funded by the National Science Foundation (FCT), under the grant SFRH/BD/147837/2019.

References

- [1] Isabelle Bos, Stephanie Vos, *et al.* The emif-ad multimodal biomarker discovery study: design, methods and cohort characteristics. *Alzheimer's research & therapy*, 10(1):64, 2018.
- [2] José Luís Oliveira, Alina Trifan, and Luís A Bastião Silva. Emif catalogue: A collaborative platform for sharing and reusing biomedical data. *International journal of medical informatics*, 126:35–45, 2019.

A Smart Mapping Editor for Standardised Data Transformation

Hannes ULRICH^{1a}, Sebastian GERMER^c, Ann-Kristin KOCK-SCHOPPENHAUER^a,
Jori KERN^b, Martin LABLANS^b and Josef INGENERF^{a,c}

^a *IT Center for Clinical Research, University of Lübeck, Germany*

^b *Federated Information Systems, German Cancer Research Center, Heidelberg, Germany*

^c *Institute of Medical Informatics, University of Lübeck, Germany*

Abstract. The integration of heterogeneous healthcare data sources is a necessary process to enable the secondary use valuable information in clinical research. Data integration is time-consuming for data stewards. The transformation using predefined rules for data harmonization can reduce the time-consuming and error-prone work and ease the data integration at various sites. In our study, we examined various script(ing) languages to find the most suitable candidate for definition of transformation rules and implement a smart editor which supports the data stewards in selecting rules reusing them. Thereby, it also provides an automatic and seamless documentation to strengthen the reliability of the defined transformation rules.

Keywords. Health Information Interoperability, Data Warehousing, HL7 FHIR

1. Introduction

The increasing digitalization of the healthcare sector produces a variety of new data sources which should be integrated to exchange data and therefore enable secondary use in clinical research [1]. But often, the origin and prior processing steps are not well documented, impeding the ability to reliably determine the data origin and thus hampering reproducible research [2]. Therefore, the data origin must be reliable. Data integration and therefore data transformation needs suitable tools to transform information with seamless documentation. Data stewards need to examine both input and output formats and define conversion rules and confirm the performed transformation. A broad variety of open source integration tools is available to extract, load and transform [3]. However, they share the same lack of exchangeability: the transformations are encoded locally in a proprietary representation. Healthcare data provider and integration centres facing the same integration problem, could benefit from exchanging these transformation schemes [4].

2. Methods

A standardized data format and a commonly used transformation language are needed to promote the exchange of transformation rules. HL7 FHIR is a suitable

¹ Corresponding author: E-mail: hannes.ulrich@itcr.uni-luebeck.de

communication standard, which defines 143 interoperable resources including three suitable for transformation: *ConceptMap*, *StructureMap*, and *Provenance*. JavaScript is one of the internet core technologies and due to the code execution in the browser on every common environment applicable.

3. Results

We took the requirements into account and implemented an editor service. The development focused a modular structure with standardized HL7 FHIR input and outputs formats, assistance for the data steward and reproducibility of the created mappings. The ConceptMap is uploaded manually or received from a FHIR server. The content is visualized with additional details about the source and target data element [5]. The editor provides syntax highlighting for various script languages and live checking to minimize errors. The data steward can define the transformation in JavaScript and use sample data to validate the mapping in real-time. A proposal function is based on previously stored transformation rules and locally-sensitive hashing function [6]. OpenID Connect was used to obtain the user information for the Provenance resource. The editor exports the resulted mapping as a FHIR bundle including a StructureMap, a Provenance and the corresponding steward in a Practitioner resource. The source code of the editor is available on <https://github.com/itcr-uni-luebeck/>.

4. Discussion

The editor focuses on standardized approach to exchange defines rules to minimize the time-consuming integration task. Our approach uses FHIR resources to promote this exchangeability and enables a *plug and play* for data integration to expedite the process. The choice of JavaScript as the transformation language ensures system-independent use since every common browser is capable of executing the rules. Nevertheless, the editor is implemented modular and language-agnostic, e.g., the use of Python is possible. In conclusion; Our study aims at closing the documentation gap for metadata-driven data integration by providing a supportive editor for the better reproducibility.

References

- [1] W.-W. Yim, A.J. Wheeler, C. Curtin, T.H. Wagner, and T. Hernandez-Boussard, Secondary use of electronic medical records for clinical research: challenges and opportunities, *Converg. Sci. Phys. Oncol.* **4** (2018) 014001. doi:10.1088/2057-1739/aaa905.
- [2] N.G. Weiskopf, and C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, *J. Am. Med. Inform. Assoc.* **20** (2013) 144–151.
- [3] T.A. Majchrzak, T. Jansen, and H. Kuchen, Efficiency evaluation of open source ETL tools, in: Proc. 2011 ACM Symp. Appl. Comput., ACM, 2011: pp. 287–294.
- [4] P. Atzeni, L. Bellomarini, P. Papotti, and R. Torlone, Meta-mappings for schema mapping reuse, *Proc. VLDB Endow.* **12** (2019) 557–569.
- [5] J. Kern, D. Tas, H. Ulrich, E.E. Schmidt, J. Ingenerf, F. Ückert, and M. Lablans, A Method to use Metadata in legacy Web Applications: The Sapply.MDR.Injector, *Stud Health Technol Inf.* (2018).
- [6] A.Z. Broder, On the resemblance and containment of documents, in: Proc. Compression Complex. Seq. 1997 Cat No 97TB100171, IEEE, 1997: pp. 21–29

Accessible Rates to Health Information on the Internet in Elderlies Increased Among Fifteen Years

Takeo SHIBATA^{a,1}, Akemi KUNIMATSU^b, Chika HIRAGA^b, and Yoko SHIMIZU^c

^a*Department of health management, Tokai University, Japan*

^b*Division of Health and Welfare, Yuzawa Town, Japan*

^c*Graduate School of Nursing, Tokyo Women's Medical University, Japan*

Abstract. Yuzawa Town, located in the Niigata prefecture of Japan, is famous for its hot springs. A citizen-centered health promotion program, Yuzawa family health plan, was initiated in 2002. It has been held for seventeen years. We evaluated changes their accessible rates to health information on the internet between 2002 and 2017 in elderlies. 431 and 435 questionnaires were corrected from elderly people at least 65 years old at 2002 and 2017. The accessible rates to health information on the internet in elderlies increased ($p < 0.001$). Profiles of accessible elderlies to health information on the internet were shown. Son and daughter might give them health information. Elderlies having any health concern or anxiety might be afraid to take health information. Accessible elderlies to health information were able to resolve suffering and breach risks by oneself.

Keywords. health promotion, health information, elderly

1. Introduction

Yuzawa Town, located in the Niigata prefecture of Japan, is famous for its hot springs. The number of citizens is about eight thousands. The population is decreasing, because Yuzawa town is one of fast aging towns in Japan. A citizen-centered health promotion program, Yuzawa family health plan, was initiated in 2002. It has been held for seventeen years. Questionnaires were collected every five years. We evaluated changes their accessible to health information on the internet between 2002 and 2017 in elderlies. We evaluated the profiles for accessible to health information on the internet at 2017.

2. Methods

431 questionnaires were corrected from elderly people at least 65 years old at 2002, and 435 questionnaires were corrected at 2017. Chi-square test was used to compare the rate of accessible health information on the internet. Chi-square test was also used to evaluate related profiles with accessible to health information on the internet at 2017 in elderlies.

¹ Corresponding Author

Evaluated profiles were living alone, living with marriage partner, living with son or daughter, living with grandchildren, living with brother or sister, health consciousness, attention to health, health concern, awakening in a morning, clockwork life, fitness habits, smoking, taking alcohol, anxiety, loneliness, consultation, resolve suffering, economic problem, handle stress, taking a rest, risk breaching, communication, life satisfaction, talking with family, job at home, intercommunion with friends, greeting with neighbors, relationship with neighbors, relationship with babies, relationship with elementary school children, relationship with junior high school students, relationship with high school students, relationship with young adults, relationship with middle aged adults, and relationship with elderlies.

3. Results

Accessible rates to health information on the internet in elderlies increased from 4.5% at 2002 to 13.1% at 2017 ($p<0.001$).

Living alone ($p<0.01$), living with marriage partner ($p<0.01$), taking alcohol ($p<0.05$), resolve suffering ($p<0.05$), risk breaching ($p<0.05$), talking with family ($p<0.01$), and relationship with middle aged adults ($p<0.01$) increased the accessible rate to health information on the internet at 2017. Living with son or daughter ($p<0.01$), health concern ($p<0.05$), anxiety ($p<0.05$), consultation ($p<0.01$), and relationships with neighbors ($p<0.05$) decreased it.

4. Conclusion

Accessible rates to health information on the internet in elderlies increased about three times between 2002 and 2017. Profiles of accessible elderlies to health information on the internet were shown. Though living alone or living with marriage partner increased accessible rates to health information, living with son or daughter decreased it. Son and daughter might give them health information. Elderlies having any health concern or anxiety might be afraid to take health information. Accessible elderlies to health information were able to resolve suffering and breach risks by oneself.

References

- [1] McKee MM, Hauser PC, Champlin S, Paasche-Orlow M, Wyse K, Cuculick J, Buis LR, Plegue M, Sen A, Feters MD. Deaf Adults' Health Literacy and Access to Health Information: Protocol for a Multicenter Mixed Methods Study. *JMIR Res Protoc*. 2019 Oct 9;8(10):e14889. doi: 10.2196/14889.
- [2] Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. *J Med Internet Res*. 2012 Jan 27;14(1):e19. doi: 10.2196/jmir.1619.
- [3] Ghweeba M, Lindenmeyer A, Shishi S, Abbas M, Waheed A, Amer S. What Predicts Online Health Information-Seeking Behavior Among Egyptian Adults? A Cross-Sectional Study. *J Med Internet Res*. 2017 Jun 22;19(6):e216. doi: 10.2196/jmir.6855.
- [4] Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. *J Med Internet Res*. 2016 Jul 18;18(7):e192. doi: 10.2196/jmir.5413.

Addressing Ethics in the CrowdHEALTH RDI Project Concerned with Large Amounts of Data to Support Health Policies

Lydia MONTANDON^{a,1}

^aResearch & Innovation, Atos Spain, Spain

Abstract. CrowdHEALTH is a R&D project involving high volumes of health-related data. Although partners have been complying with all regulatory and ethical issues according to their context in different EU countries, ethical aspects related to the adoption of the project results and possible impact on individuals and populations were not in the project agenda from the outset. Inspired by an increasing number of initiatives in the EU concerned with aspects related to ethics in research and innovation, future plans consider investigating the level of awareness of all stakeholders involved in the project about the potential implications of their actions in terms of ethical acceptance of project outcomes.

Keywords. Ethics, Ethical Issues, Responsible Innovation, RRI

1. Introduction

CrowdHEALTH is an EU research project that intends to collect and aggregate high volumes health-related heterogeneous data from multiple sources with the aim of supporting policy making decisions. A Policy Development Tool proposes an interface that allows policy makers to interact with data and key performance indicators to create visual representations that support the elaboration or evaluation of Health policies.

At project start, mainly ethical issues that may affect the research and the testing of developed results in real cases pilots were considered. In the context of CrowdHEALTH, ethical issues were principally centered on privacy and data protection and children and patients' involvement.

2. Method

The project collected information related to ethical requirements from partners that had to run pilots in their organizations. A structured template with hints for its fulfillment was provided to gather consistent data from all use cases. Contributions were integrated, extracting general procedures that applied to all pilots and describing the procedures specific to each pilot.

¹ Corresponding Author, E-mail: lydia.montandon@atos.net.

3. Results

The compiled document showed that all measures to ensure that the project complied with ethical principles and relevant national and EU legislation had been taken.

4. Discussion

However, considering CrowdHEALTH's intention is to facilitate the access and analysis of large amounts of data for policy – and maybe intervention in the future – that may have an impact on whole populations, potential ethical implications go beyond mere research ethics. The question arises whether the project is dealing with all ethical aspects in a proper way. The project faces ethical challenges related to subjecting individuals and populations to policies and interventions based on non-transparent data sources, processing and analyses. Although CrowdHEALTH is developing components to clean data and ensure quality and reliability, it is not fully clear who will be accountable for the quality of the data input, or the algorithms applied.

Inspired on the one hand by recent concerns and interest in the challenges that ethical aspects present when using technologies related to Big Data and Artificial Intelligence, where transparency and accountability are treated as key issues and, on the other hand by the EU's Responsible Research and Innovation (RRI) approach, implying that research in Europe needs to understand, anticipate, and address ethical, legal and social aspects, the project started wondering if all experts involved in the project were aware and applying methods to ensure that the results of the project are desirable and anticipate risks. We therefore started to shift our attention towards what could be done to anticipate effects related to ethical issues that would impact the adoption of the project outcomes.

5. Conclusive Remarks

It becomes evident that the project needs to investigate more if there is awareness within the team and if an individual questioning of choices and actions would be enough to ensure that project consortium culture is aligned with the EU RRI approach.

To obtain insight about all stakeholders involved in the project (researchers, designers, developers, etc.) awareness on ethical issues in their daily work, a survey is being launched, including a few demographic questions, to allow clustering per profile.

The objective is to reach directly individual experts along the research and innovation value chain, and understand how familiar they are with ethical aspects, such as transparency, accountability, social values, etc. and if this may result as a better understanding of ethical implication of the project outcomes on society.

Acknowledgements. The CrowdHEALTH project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727560.

Advanced Machine Learning in Prediction of Second Primary Cancer in Colorectal Cancer

Chi-Chang CHANG^a, Ying-Chen CHEN^a

^a *Department of Medical Informatics, Chung Shan Medical University & IT Office, Chung Shan Medical University Hospital, Taichung, Taiwan*

Abstract. Colorectal cancer (CRC) ranked third among most commonly diagnosed cancers worldwide. The onset of second primary cancer (SPC) is an important indicator in treating CRC. We tried to use the advanced machine learning method in order to find the factors of SPC. Patients with CRC from three medical centers were identified from cancer registries in Taiwan. The classifier of A Library for Support Vector Machines (LIBSVM) and Reduced Error Pruning Tree (REPTree) were applied to analyze the relationship of clinical features with category by constructing the optimized model of every classified issue. Machine learning can be used to rank the factor affecting the secondary primary malignancy. In the clinical practice, physician should be aware the possibility of cancer recurrence and routine checkups for early second primary malignancy detection is recommended. The accuracy rate of the may need more big data. The machine learning method is feasible in detecting/predicting potential second primary cancer in the future.

Keywords. Colorectal Cancer (CRC), Second Primary Cancer (SPC), Machine Learning method.

1. Introduction

Colorectal cancer (CRC) is the third most common cancer in the world and the fourth most common cause of death overall [1]. It is also the most common type of cancer in Taiwan, the incidence of CRC is approximately 16%. Early detection of second primary cancers is essential to improve health outcomes of cancer patients. Therefore, a valid and reliable method/tool to help detecting the SPC is necessary.

2. Methods

Our study identified data associated with three medical centers from the cancer registry database (Taiwan). We used the ICD number C18–C20 to select the CRC patients. The colorectal cancer was all diagnosed from 2004 to 2012 and the age was 18 years old and above. A total of 4299 patients with primary colorectal cancer were enrolled. Of those, 541 samples had primary colorectal cancer plus second primary malignancies. In this study, we used the correlation analysis to examine the following 20 factors associated with colorectal cancer: patient age, primary site, histology, behavior code, differentiation, tumor size, pathologic stage, surgical margins, surgical procedure,

radiation therapy, pre-operative radiation therapy, regional body order, highest dose, maximum number of times, lowest dose, lower number of times, body mass index (BMI), smoking, areca and drink. We use LIBSVM (A Library for Support Vector Machines)[2] with the radial basis function (RBF) kernel to construct predictive models and optimize the C and γ parameters of each model. *REPTree Reduced Error Pruning Tree (REPTree)* uses the regression tree logic and creates multiple trees in different iterations. The predictive ability of each system was evaluated using accuracy (Acc), sensitivity (Sn), specificity (Sp), and the Matthews correlation coefficient (MCC).

3. Results

By LIBSVM implementation, the model building employed 8 features, which includes behavior code, differentiation, regional body order, patient age, areca, surgery, radiation therapy and lowest dose, could achieve the equivalent performance outcome with the model using all features in the analysis. Applying the REPTree algorithm, we found that the decision tree employed three features of patient age, differentiation and organizational patterns have better accuracy of MCC in 0.294 and Sn in 0.706 than original model by parameters optimization. From the decision tree architecture, we discovered the value of differentiation > 9.5 was classified as SPM, when value of differentiation < 9.5 , the organizational patterns and patient age were as the leaf node to classify the SPM and NSPM.

4. Discussion

Our study has explored risk factors for predicting second primary malignancy and discovered the top four important features were pStage, surgical margins, smoking and drinking. In addition, we found that the REPTree algorithm used only the three features (patient age, organizational patterns and differentiation) to optimize the parameters of REPTree model to improve performance. With machine learning program, we have developed a feasible and a robust method to identify four important factors: pStage, surgical margin, smoking and drinking contributing to the second primary cancer and recurrence in colorectal cancer. Mechanical learning can be used as an effective medical decision-making tool to assist physicians to establish a prognosis and improve their diagnostic accuracy. We strongly recommend clinicians to consider using mechanical learning method in diagnosing and treating cancer patients to provide high quality care.

References

- [1] Global Burden of Disease Cancer, C., et al., *Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study*. JAMA Oncol, **2018**.
- [2] C.-C. Chang, and C.-J. Lin, *LIBSVM: a library for support vector machines*. ACM transactions on intelligent systems and technology (TIST), **2(3)** (2011), 27.

Agent-Based Modeling and Simulation of Care Delivery for Patients with Thrombotic and Bleeding Disorders

Noemi GIORDANO^a, Samanta ROSATI^a, Federica VALERI^b, Alessandra BORCHIELLINI^b and Gabriella BALESTRA^{a,1}

^a*Department of Electronics and Telecommunications, Politecnico di Torino, Italy*

^b*Oncologic Department, Città della Salute e della Scienza di Torino, Torino, Italy*

Abstract. The quality of patients care delivery is thought to be strongly affected by the physicians' workload. In this study we present an Agent-Based model of the processes during a typical working day. We simulated the current scenario and a possible scenario concerning the introduction of a second ambulatory as a potential improvement in the center organization. Our results validated the reliability of the model and showed that the introduction of a second ambulatory averagely reduces the daily physician' workload.

Keywords. ABM, simulation, process modelling, clinical process, NetLogo.

1. Introduction

It is nowadays well known that the delivery of high-quality healthcare services to patients and the resulting clinical outcomes are dependent on the workload of the clinical staff. Physician's workload higher than normal and consequent stress were proven to negatively impact on clinical outcome and on the quality of care [1]. Computational modelling and simulation is a unique tool to get a deeper insight in complex systems' behavior and to predict the effect of changes [2]. The diffusion of Agent-Based Modelling (ABM) strongly increased across the latest decades to face the growing complexity of the world [2]. The aim of this work is to present an ABM of the care processes for patients with thrombotic and bleeding disorders with different scenarios and their effects on the physician workload. The study was conducted at the "Malattie Trombotiche ed Emorragiche" center of a Turin hospital.

2. Materials and methods

In a previous work [3], we described the process involved and performed data collection. Two Process Modelling tools, namely Synopsis diagrams and Swim lane activity diagrams, were used to analyze all clinical processes carried out by the two

¹ Corresponding Author: Gabriella Balestra, E-mail: gabriella.balestra@polito.it.

physicians of the center during a typical working day. The results were used to perform the ABM modeling and simulation. Among available ABM software and tools, we chose NetLogo 6.1.0 platform¹ for its simplicity in use, high scalability and high computational modelling strength.

In our model, we defined three types of agents: 1) “**Physician**”: specialists of the center; 2) “**Ambulatory patient**”: patients with scheduled ambulatory visit; 3) “**Internal advice patient**”: patients needing internal advice. Agents correspond to the actors previously identified as involved in clinical processes [3]. All agents moved and performed the activities in the simulation environment that was defined accordingly to the scaled map of real hospital. The model rules were defined by implementing the processes as previously modelled [3]. The simulation was based on several parameters statistically defined based on the collected data [3]. First, a simulation of an average typical day in the present situation was carried out to validate the reliability of the model. Then, since the ambulatory care absorbed the highest percentage of the workload, a new scenario was simulated to evaluate the effect of the availability of a second ambulatory room. Two different situations were modelled: (a) the second room was used only when ambulatory visits were performed both in the morning and in the afternoon; (b) the additional ambulatory was used in parallel with the first one.

3. Results and Conclusions

The results of the simulations are compared with the collected data. The overall physicians’ workload (expressed in minutes) was 557 ± 74 for the current scenario, 561 ± 76 for the new scenario (a), and 555 ± 58 for the new scenario (b). Comparing the collected data (566 ± 135 minutes) with those obtained from the simulation of current scenario it can be noticed that the results are consistent demonstrating the accuracy reliability of the developed model. Moreover, it can be observed that the introduction of a second ambulatory room reduces the physician workload. Process modeling and ABM proved to be suitable tools to simulate clinical processes and to provide insight on possible organizational changes.

References

- [1] P. van den Hombergh, et al., High workload and job stress are associated with lower practice performance in general practice: an observational study in 239 general practices in the Netherlands, *BMC Health Serv. Res.* **9** (2009) 118.
- [2] D. Fone, et al., Systematic review of the use and value of computer simulation modelling in population health and health care delivery, *J. Public Health (Bangkok)*. **25** (2003) 325–335.
- [3] S. Rosati, et al., Characterization of Physicians Workload in a Reference Center for the Treatment of Thrombotic and Bleeding Disorders *, in: Proc. 2019 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE, 2019: pp. 1359–1362.

¹ U. Wilensky, NetLogo, Cent. Connect. Learn. Comput. Model. Northwest. Univ. Evanst. (1999)

An “in silico” Bench to Bedside Approach to Investigating Sepsis Biomarkers

Grace O'MAHONY^a, Christopher HAWTHORNE^a, Karl MCQUAID^a and Guillermo H. LOPEZ-CAMPOS^{a,1}

^a*Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, UK*

Abstract. Sepsis results in various patient complications and is due to a heightened inflammatory response against infection. This condition requires further exploration of biomarkers. We employed an “in silico” method comprised of text-mining and additionally clinical validation through the Medical Information Mart for Intensive Care. We highlight that Factor VIII shows potential as a pertinent septic shock biomarker..

Keywords. Biomarkers, Bioinformatics, Sepsis, Natural language processing

1. Introduction

Biomedical research is producing new data and knowledge at an unprecedented pace and is through an increasingly vast array of databases and data repositories containing a large variety of data contents and formats. This is leading to new avenues of research where it is possible to construct new experimental designs without needing to generate new data instead using the already existing published biomedical data.

Sepsis is a condition caused by an overwhelming widespread inflammatory response to infection resulting in massive deregulation of the immune response. Globally, sepsis occurs in approximately 10 per 1000 hospital patients, of which 30% of these patients develop multiple organ syndrome, 60-80% of progression to septic shock and mortality rates reaching 20% [1].

In this work we propose an “in silico” bench to bedside approach to analyse free-text data (also typically referred as “unstructured data”) extracted from publicly available resources which might facilitate the identification of sepsis biomarkers. The aims of this project are to identify molecular biomarkers using text mining approaches and then validating them in a clinical dataset combining different medical coding systems to gain deeper insights into this condition.

2. Methods

The initial stage of the project involved identifying relevant biomarkers for sepsis. For this purpose, we use an in-house developed methodology [2] that combines PubMed

¹ Corresponding Author,; E-mail: g.lopezcampos@qub.ac.uk.

queries and Pubtator annotations [3] using R. This approach enables us to retrieve chemical and other protein/gene biomarkers.

For clinical validation and analysis of the identified chemical biomarkers (encoded as MeSH or ChEBI terms) we used data from MIMIC-III, the 'Medical Information Mart for Intensive Care' [4], an openly available dataset comprising of de-identified health data collected from 61,532 intensive care unit admissions (53,432 adult patients stays and 8,100 neonatal), spanning from June 2001 – October 2012. Data collected from all aspects of ICU patient care includes patient details, vital signs, laboratory tests using LOINC coding and ICD-9 codes.

Relevant patients were selected using ICD-9 codes for sepsis, severe sepsis and septic shock and a set of non-septic controls involving different degree of inflammation from epilepsy (EPI) and ventilator associated pneumonia (VAP) diagnosis.

3. Results

5,566 chemicals, including many antibiotics and proteins, were retrieved from the initial analyses of the literature abstracts. Nine out of the top 20 occurring (as frequency in abstracts) potential sepsis biomarkers could be mapped to related LOINC codes. Those 9 codes were subsequently searched in MIMIC-III to check that they were present in the groups of patients of interest with sufficient sample sizes for statistical analysis. The LOINC code for Factor VIII fulfilled all the previous criteria and was chosen for analysis comparing sepsis samples and controls. We performed Mann-Whitney unpaired tests which yielded the following results "Sepsis vs EPI" $p=0.5714$; "Severe Sepsis vs EPI" $p=0.3928$; "Septic shock vs EPI" $p=0.04843$ and "Septic shock vs VAP" $p=0.04492$.

4. Conclusions

These results show that Factor VIII could be a potential element of interest to distinguish septic shock from controls and would be coherent with the development of coagulopathy in sepsis.

This still under development approach aims to integrate and automate an "in silico" bench-to-bedside platform for the analysis and eventually the validation of biomarkers in sepsis. This approach could be extended to other areas of interest in intensive care.

References

- [1] G. Polat, R.A. Ugan, E. Cadirci, Z. Halici, Sepsis and Septic Shock: Current Treatment Strategies and New Approaches, *The Eurasian Journal of Medicine* **49** (2017), 53-58.
- [2] G. Lopez-Campos, E. Bonner, L. McClements. An Integrative Biomedical Informatics Approach to Elucidate the Similarities Between Pre-Eclampsia and Hypertension. *Stud Health Technol Inform.* **264** (2019), 264:988-992.
- [3] C.H. Wei, H.Y. Kao, Z. Lu, PubTator: a web-based text mining tool for assisting biocuration, *Nucleic acids research* **41** (2013), 518-522.
- [4] A.E. Johnson et al., MIMIC-III, a freely accessible critical care database, *Scientific Data* **3** (2016), 160035.

An Online Prevention Tool for Ophthalmological Disorders

Maria PATTE¹, Joseph LIASKOS, Parisis GALLOS and John MANTAS
Health Informatics Laboratory, School of Health Sciences, National and Kapodistrian University of Athens, Greece

Abstract. The information of the general population regarding the prevention of eye diseases is crucial. This work presents the design and development of an online prevention tool for the information and education of the general population about ophthalmological diseases. A literature review was conducted for the collection of all necessary and up-to-date information regarding the diseases selected. The phases of requirement analysis, design, development and evaluation were followed. The online tool developed, proved to accomplish the aims of usability and usefulness in the prevention of eye diseases and their complications.

Keywords. Ophthalmological Diseases, Prevention, Primary Healthcare

1. Introduction

The regular eye health monitoring plays a crucial role in the prevention and correct treatment of ophthalmological diseases [1]. Despite of the common belief that ophthalmological diseases are more frequent to elders, there are some diseases which appear to younger people and evolve over time [2]. Since an early diagnosis and treatment can prevent the loss of vision and other complications, it is necessary to educate the general population and raise its awareness of eye health [3,4]. The purpose of this work is to present the development of an online prevention tool which informs the general population about ophthalmological diseases.

2. Methods

The pilot online tool for the prevention of the ophthalmological diseases followed the phases of requirement analysis, design, development and evaluation. In the requirement analysis the functions of the tool were defined, and a literature review was conducted to collect all the necessary and up-to-date information regarding these diseases. 18 publications were used to structure the content of the tool. The criteria for the selection of the eye diseases were that they are very common and serious, they could lead to blindness if remained untreated, and their prevention is necessary and relatively easy. Based on the requirement analysis, an educational and instructional online tool was designed in order to serve the principles of usability, user-friendliness and usefulness. The language used had to be simple for the general population. A site map was structured,

¹Corresponding Author.; E-mail: mary_patte@hotmail.com.

and the most appropriate modules were chosen, according to the potential users' needs and the requested functions of the website. The Content Management System (CMS) technology was used in the development phase. Finally, an evaluation was conducted to assess the users' opinion regarding the usability and usefulness of the tool through semi-structured interviews. Three small focus groups were composed and the criteria of selecting the participants were their age, and their familiarity with technology or healthcare sector.

3. Results and Discussion

A pilot online prevention tool (for primary and blindness prevention) was developed in Greek language targeting the general population. The content includes informational and educational material concerning the anatomy of the eye, some of the most frequent eye diseases, specific eye exams, articles on various eye-related topics, as well as, advises and guidelines about eye diseases. Some examples of the eye diseases included are cataract, glaucoma, strabismus, age-related macular degeneration (AMD) and diabetic retinopathy. The users can perform a self-examination test for age-related macular degeneration. In addition, the user can request for more information or advice from experts through a form. A main menu in the site allows the navigation to all information and functions. The evaluation results concerning usability and usefulness were positive. All evaluators agreed that the website is easy to understand and navigate. They noted that the presentation of the information is simple, adequate, with no confusion, promoting education and prevention; they agreed that this tool promotes the awareness regarding the Ophthalmological Disorders. Healthcare professionals also mentioned that this tool can be useful for the clinical practice.

4. Conclusions

The described online tool targeting the general population contributes to the prevention of ophthalmological diseases in a simple and easy-to-use way. Compared to similar websites it includes related content. The inclusion of more ophthalmological diseases (e.g. refractive errors of the eye), audiovisual material, and "search" functionality for visiting ophthalmologists, are some of the limitations defined. Future work could include more content and interactivity with the user, as well as the usage of this tool on a larger scale.

References

- [1] National Academies of Sciences, Engineering, and Medicine, *Making eye health a population health imperative: Vision for tomorrow*, National Academies Press, 2017.
- [2] World Health Organization, Universal eye health: a global action plan 2014-2019, 2013.
- [3] J. Ramke, A.B. Zwi, A. Palagyi, I. Blignault, and C.E. Gilbert, Equity and blindness: closing evidence gaps to support universal eye health, *Ophthalmic epidemiology* **22**(5) (2015), 297-307.
- [4] J. Beynat, A.B. Mehidi, J.P. Aubert, A.M. Bron, P. Massin, and C. Creuzot-Garcher, RETIDIAB®: Assessment of a continuing medical education website for the improvement of diabetic retinopathy management, *Diabetes & metabolism* **37**(2) (2011), 118-123.

Analysis of Professional Competencies for the Clinical Research Data Management Profession

Saly ABOUELENEIN^{a,1}, Tremaine WILLIAMS^a, Jaime BALDNER^a and
Meredith Nahm ZOZUS^{a, b}

^aUniversity of Arkansas for Medical Sciences, Little Rock Arkansas, USA

^bUniversity of Texas at Health, San Antonio Texas, USA

Abstract. Objective: This job analysis was conducted to compare, assess and refine the competencies of the clinical research data management profession.

Materials and Methods: Two questionnaires were administered in 2015 and 2018 to collect information from data managers on professional competencies, types of data managed, types of studies supported, and necessary foundational knowledge.

Results: In 2018 survey, 67 professional competencies were identified. Job tasks differed between early- to mid-career and mid- to late-career practitioners. A large variation in the types of studies conducted and variation in the data managed by the participants was observed.

Discussion: Clinical research data managers managed different types of data with variety of research settings, which indicated a need for training in methods and concepts that could be applied across therapeutic areas and types of data.

Conclusion: The competency survey reported here serves as the foundation for the upcoming revision of the Certified Clinical Data Manager (CCDMTM) exam.

Keywords. Professional competencies, clinical data management, SCDM

1. Introduction

The Society for Clinical Data Management (SCDM) is a non-profit organization for Clinical Data Management professionals. It published the first version of the Good Clinical Data Management Practices (GCDMP) in 2000 [1], and established the Certified Clinical Data Manager (CCDMTM) program in 2004. The job tasks of professional clinical data managers have periodically been surveyed to maintain the exam [2].

2. Methods

This survey was administered at the SCDM meeting, 23-26 September 2018, in Seattle WA. The tasks were obtained from previous surveys and job analysis work through the SCDM [2, 3].

¹ Corresponding Author.; E-mail: sabouelenein@uams.edu

3. Results

The response rate for our survey was 58.8%. The majority of the respondents 77.6% worked in private or publicly traded company. Respondents indicated a variety of organization types, the most frequently reported being pharmaceutical companies 38.7%. Eighty-two percent of the respondents indicated that they were practicing clinical data managers and 55% of the respondents indicated that they supervised clinical data managers. Respondents reported an average of 12.6 years working in clinical data management, and 16% of them reported being certified.

Most respondents (90.8%) managed data for randomized controlled trials (RCTs), but respondents also managed data for other types of studies. Respondents reported clinical data management tasks performed in their organization across the spectrum of design, programming, data processing, testing, project management, and training and review task domains. Future tasks they expected to perform included, risk-based monitoring, centralized monitoring, data visualization, maintaining awareness of new technology, new technology adoption, process improvement, metadata management, and data warehousing. At 77% and 74%, the respondents indicated data from personal wearable devices and data from home monitoring devices as types of data they expected to manage in the future.

4. Discussion

Our survey updated the previous professional job analysis and helps to monitor changes in work done by professional Clinical Research Data Managers. Changes between the 2015 and 2018 data confirmed a trend toward growth of the profession in terms to managing data for varied types of studies and data. This stands in contrast to the historical norms of data managers serving mainly one type of research and a limited variety of data. New tasks are consistent with application of basic informatics methods in clinical research contexts. This diversification calls for curricula and formal degree programs to educate Data Managers in general principles and methods that can be applied to new research settings, studies across the continuum of clinical research and varied types and sources of data.

5. Conclusion

The results of this study serve as the foundation for the Certified Clinical Data ManagerTM exam and inform the scope of the practice standards for the profession, the GCDMP.

References

- [1] Society for Clinical Data Management, Good Clinical Data Management Practices (GCDMP), October 2013. Available from <http://www.scdm.org>
- [2] .Zozus MN, .Lazarov A., Smith LR. et al. Analysis of professional competencies for the clinical research data management profession: implications for training and professional certification. *J Am Med Inform Assoc.* **24**(2017), 737–745.
- [3] Bornstein S. Clinical data management task list. *Data Basics* (1999); 5:8-10.

Application of IBM Watson to Support Literature Reviews: A Preliminary Experience in Restorative Dentistry

Giacomo DERCHI^a, Marco VISENTIN^b, Vincenzo MARCHIO^a, Lisa LARDANI^a, Antonio BARONE^a, Marco PRENASSI^b, and Sara MARCEGLIA^{b,1}

^aUniversità degli Studi di Pisa, Pisa Italy

^bUniversità degli Studi di Trieste, Trieste, Italy

Abstract. Literature reviews are crucial in the choice of the best personalized material type and restauration type in restorative dentistry. We developed an IBM-Watson based system to support literature search for restorative dentistry, and compared its results to a literature search performed by a trained professional. We found that our system could assist the researcher in performing a literature review, but the grounding semantic model needs to be refined in order to provide more extensive results.

Keywords. Literature review, restorative dentistry, cognitive computing.

1. Introduction

Technology for restorative dentistry is becoming more and more advanced, and new techniques and materials have been introduced to obtain minimally invasive treatments and better outcomes[1]. When choosing the best solution for patients, clinicians should rely on the available growing evidence. In this scenario, a system supporting the process of literature review[2] will optimize the choice of materials for restorations, and, in turn, the patient's quality of life. To this end, we tested the ability of IBM Watson cognitive computing services to provide the ground engine for a system supporting dental inlays literature search process.

2. Methods

PubMed API were used to run the literature search. A concept model was developed and trained in IBM Watson Knowledge Studio using 42 annotated documents (review papers on dental inlays). The service uses proprietary natural language processing methodologies. The precision of the system (percentage of true positives on the overall positives recognized by the model) and the recall (percentage of true positives on the overall true from the ground truth) were used as performance measures of the model. The model was then implemented in IBM Watson Discovery in order to process PubMed abstracts and to retrieve the concepts, relationships, and specifications according to the semantic model. A user interface developed using Node-Red[3] allows the clinician to visualize the sentences (and abstracts) more relevant to the research question related to the restauration type, the material type, and the characteristics. The

¹ Corresponding Author,; E-mail: smarceglia@units.it.

system was validated comparing the literature review automatically performed with the literature review performed by a trained clinician, in terms of number and relevance of papers retrieved.

3. Results

The semantic model included 15 main entities (among them, “restauration type”, “material type”, “character”, “comparator”, “value”) and 13 main relationships and achieved an overall precision of 93% and an overall recall of 80%. In Watson Discovery, a total of 165 documents (abstracts) were processed using the general PubMed query. The system was tested with the aim to identify the papers relevant to establish the longevity comparison between composite inlays and ceramic inlays. The first man-made search yielded a total of 24 unique papers, 8 of which were included. The first search made by IBM Watson yielded a total of 18 unique papers, 9 of which were included. The two searches obtained completely different results, with only 3 papers found by both the human researcher and IBM Watson, but all of them potentially fit to be included in a literature review.

4. Conclusions

These preliminary results show that IBM Watson could be used to assist the researcher in performing a literature review, but the model on which Watson is trained needs to be refined in order to provide more extensive results. These findings will help to further refine IBM Watson's abilities and getting closer to obtaining a Cognitive Computing Platform able to answer clinical questions made by Dental professionals

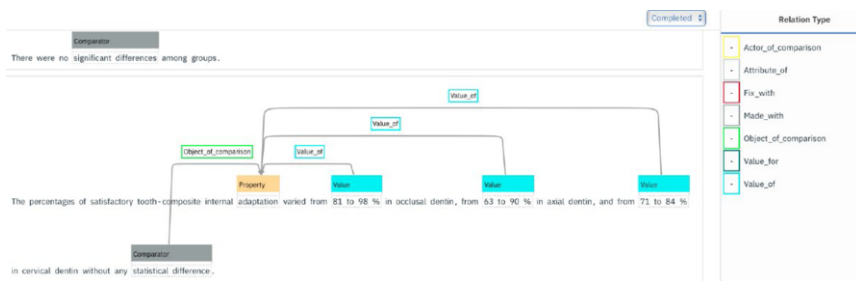


Figure 1. Text annotation in Watson Knowledge Studio with the defined semantic model.

References

- [1] Fron Chabouis H., Smail Faugeron V., Attal J.P. Clinical efficacy of composite versus ceramic inlays and onlays: A systematic review. *Dental materials*. **29** (2013), 1209-18.
- [2] David Ferrucci, Anthony Levas, Sugato Bagchi, David Gondek, Erik T. Mueller. Watson: Beyond Jeopardy!, *Artificial Intelligence*, Volumes **199–200**, (2013), 93-105.
- [3] Node Red for Watson: <https://flows.nodered.org/node/node-red-node-watson>, accessed on Jan 14, 2020.

The Impact of Information Quality on Retracted Bioinformatics Literature

Austin SPRINGER^a, Donna CHACHERE^a, Awaad ALSARKHI^a, and Meredith ZOZUS^a

^aUniversity of Arkansas at Little Rock

Abstract. One of the biggest challenges facing biomedical research today is the lack of reproducibility in findings. In response, a growing body of literature has emerged to address this. However, much of this focuses on bias and methods, while little addresses the issue of information quality. The purpose of this poster is to determine the role of information quality for retracted bioinformatics literature.

1. Introduction

Irreproducibility continues to be one of the more emerging issues in recent biomedical research. Much has been written about its negative impact on the progression of biomedical science and ultimately patients [1], with guidelines proposed [2,3]. Few of these, however, address the impact of low information quality (IQ). In this poster, the IQ issues found in retractions are discussed, and corresponding patterns analyzed.

2. Method

The Retraction Watch database [4] was filtered by: Retraction or other Notices From Date value 1/1/2018, To value 12/31/2018, and Reason(s) for Retraction value “Concerns/Issues About Data” which resulted in 124 retractions with fields such as Subject, Journal, Publisher, Original Paper Date, Retraction date, and Country. Of these, 119 were classified for associated IQ dimension by 2 independent coders. After obtaining the Scimago Journal Rank of each record, statistical analysis was then done.

3. Results

54% of retractions were due to the Believability dimension of IQ. China was the most prevalent country being involved in 32% of all records, followed by India 24%, and the U.S. 19%. SJR score comparisons indicate records involving the top 2 IQ reasons (Believability & Wrong data used) are associated with higher-ranking journals ($p = 0.05$), as well as significant differences between regions ($p = 0.02$). Articles published in the U.S. & Canada through publishers other than Elsevier, PLoS & Springer were 3.3 times less likely to be associated with Believability than all other articles ($p = 0.05$). Survival analysis indicates studies having Biochemistry as a subject have a half-life of 33.8 months and remain in publication longer than those with other subjects.

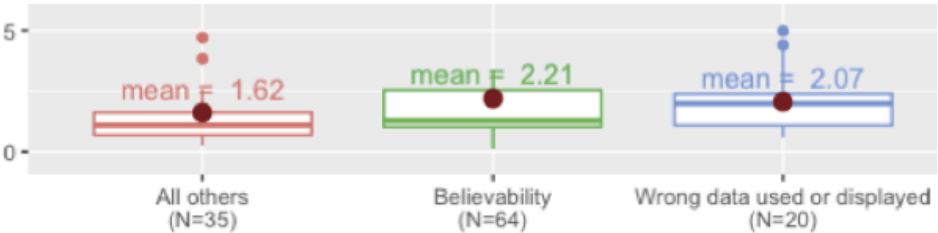


Figure 1. the association of higher journal scores with the top 2 IQ issues.

Table 1. Articles least associated with Believability

	Believability	All other issues
U.S. & Canada articles not published by Elsevier, PLoS & Springer	4	10
All other countries & publishers	60	45

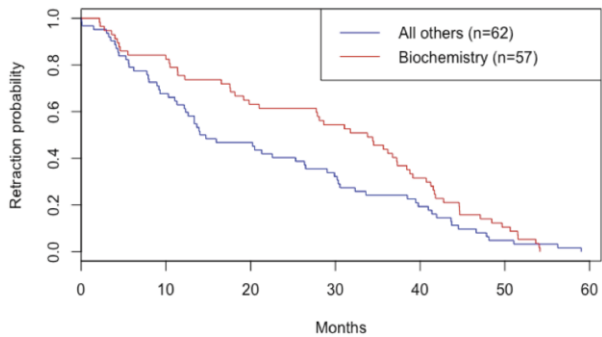


Figure 2. survival analysis showing how those with Biochemistry subject remain in publication longer.

4. Discussion & Conclusion

Our aim was to investigate what role IQ plays for retracted bioinformatics studies. While Believability is the most prevalent, others are associated with countries, subjects, and longevity in months published. Future work could involve expanding to a wider date range, as our study design was limited to calendar year 2018.

References

[1] Mullane, K. and M. Williams, Enhancing reproducibility: Failures from Reproducibility Initiatives underline core challenges. *Biochemical Pharmacology*, 2017. 138: p. 7-18.

[2] Begley, C.G.J.N., Reproducibility: Six red flags for suspect work. 2013. 497(7450): p. 433.

[3] Jarvis, M.F. and M. Williams, Irreproducibility in Preclinical Biomedical Research: Perceptions, Uncertainties, and Knowledge Gaps. *Trends in Pharmacological Sciences*, 2016. 37(4): p. 290-302.

[4] Watch, R. 3/1/2019]; Available from: <http://retractiondatabase.org/RetractionSearch.aspx>.

Association of C>U RNA Editing with Human Disease Variants

Sapan MANDLOI^{a1}, Zackary FALLS^a, Rong DENG^b, Ram SAMUDRALA^a, Peter L. ELKIN^{ac}

^a Department of Biomedical Informatics, University at Buffalo, Buffalo, NY, USA

^b Jacobs School Of Medicine And Biomedical Sciences, Buffalo, NY, USA

^c Department of Veterans Affairs, VA Western New York Healthcare System, Buffalo, NY, USA

Abstract. RNA-editing is an important post-transcriptional RNA sequence modification performed by two catalytic enzymes, "ADAR"(A>I) and "APOBEC"(C>U). Although APOBEC-mediated C>U editing has been associated with a number of human cancers, the extent of C>U editing in human disease remains unclear. Here, we performed an association study and found that at least 1293 human disease variants occur at sites predicted by sequence motif analysis (RNASee protocol) to undergo APOBEC3A/G C>U editing. These variants were associated with a wide array of human disease conditions ranging from cancer, metabolic disorders, retinopathies, cardiomyopathies, neurodegenerative disorders and immunodeficiencies. These results indicate that APOBEC mediated C>U RNA editing may have widespread and previously unreported contributions to human disease conditions..

Keywords: RNA-editing, Protein diversity, Cytidine deaminase

1. Introduction

RNA-editing is an important post-transcriptional RNA sequence modification performed by two catalytic enzymes, "ADAR"(A>I) and "APOBEC"(C>U)¹. An emerging body of research has implicated APOBECs playing important roles in human disease processes such as cancer^{2,3}, Alzheimer disease and primary pulmonary hypertension⁴. Although A>I RNA editing has been widely studied, the extent of C>U editing in human disease processes is still largely unclear, and no comprehensive database yet exists for C>U editing events. Using the features of the experimentally verified RNA editing sites, we have developed a program to predict editing sites and compared with existing gene-disease-variant information to further assess the extent to which C>U editing may contribute to human disease.

2. Methods

2.1 Predicting C>U RNA editing sites and data collection

The RNASee platform (<http://128.205.176.246/masee/>) identifies tetra-loop sequences by scanning the mRNA sequence (12789 CDS fasta sequence collected using UniProt REST services) with a four nucleotide loop (0-3) sliding window in order to identify target sequences containing a CC at the 2-3 positions. The secondary structure of the sequence containing putative C>U editing site for APOBEC3A and 3G is then predicted using ViennaRNA (<https://www.tbi.univie.ac.at/RNA/>) and the resulting energy is used to support the structural stability. We compared the RNAsee list to the humsavar human variant list from UniProtKB/Swiss-Prot. We found 30705 (39%),

¹ Corresponding author; Email: sapanman@buffalo.edu

40035 (51%), 8081 (10%) total variants are classified as human disease variants, non-disease associated variants (polymorphism), and indeterminate variants respectively. We identified 1278 functional sites from UniProt-SwissProt database (from 206/734 genes/proteins), consisting of 9% active sites, 19% binding sites, 54% glycosylation sites and 18% metal-ion binding sites.

3. Results

3.1 Predicted RNA editing sites are associated with known human disease protein variants and pathway analysis

RNAseq generated a list of TOP25 ranked probable C>U RNA edit sites, associated with 78852 human variants and 12789 genes. A comparison of C>U RNA editing sites with known human-variants found 1293 out of 30705 (4%) variants and 734 genes were associated with known human disease processes, 1828 out of 40035 (5%) variants were associated with polymorphisms and 191 out of 8081 (2%) variants were unclassified. The top 5 most frequently diseases and genes associated with C>U variants are listed in Table 1. Pathway analysis showed enrichment of angiogenic, hematologic and hypoxia-induced processes, consistent with recent in-vitro studies⁵. In addition, we identified functional sites of metabolically essential proteins, including FGFR1, GALNS, G6PC and FTO, suggesting these editing events may have direct impact on protein function.

Table 1: Top 5 associated diseases and genes

Genes(# associated diseases)	Disease (# associated genes)
TP53 (6)	Pheochromocytoma (9)
VHL (6)	Dejerine-Sottas syndrome (6)
CDKN2A (5)	Diabetes mellitus, permanent neonatal (6)
G6PC (5)	Li-Fraumeni syndrome (6)
FGFR1 (5)	Sporadic cancers (6)

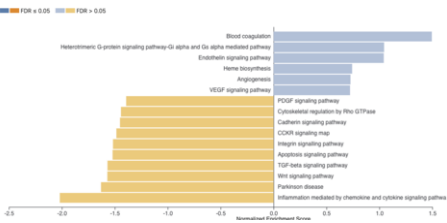


Figure 1: List of pathways

4. Conclusions

In this study, we were able to predict a large number of previously unidentified APOBEC mediated C>U editing sites across the human transcriptome and associate them with a wide array of human disease conditions. Our results suggest that APOBEC-mediated C>U editing may play a key role in hypoxia-induced remodeling during oncogenesis. In future research will apply our analysis to the transcriptomes/proteomes of specific disease conditions in order to further clarify the functional and predictive role of APOBEC mediated C>U RNA editing in human disease.

References

[1] Tang W, Fei Y, Page M. Biological significance of RNA editing in cells. Mol Biotechnol. 2012;52(1):91-100.

[2] Baysal BE, Sharma S, Hashemikhabir S, Janga SC. RNA Editing in Pathogenesis of Cancer. Cancer Res. 2017;77(14):3733-9.

[3] Sharma S, Patnaik SK, Kemer Z, Baysal BE. Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes. RNA Biol. 2017;14(5):603-10.

[4] Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet. 2015;47(9):1067-72.

[5] Sharma S, Wang J, Alqassim E, Portwood S, Cortes Gomez E, Maguire O, et al. Mitochondrial hypoxic stress induces widespread RNA editing by APOBEC3G in natural killer cells. Genome Biol. 2019;20(1):37.

Acknowledgements

This work has been supported in part by grants from NIH NLM T15LM012595, NIAA R21AA026954, and NCATS UL1TR001412. This study was funded in part by the Department of Veterans Affairs.

Automated Analysis of the Heterogeneity of Histological Glioblastoma Slides Using Neural Networks

Michael OERTL^{a,b,1}, Georg PROKOP^c, Felix HOLL^{a,d,e}, Marina FOTTELER^{a,f}, Viktoria MUEHLBAUER^a, Walter SWOBODA^a and Friederike LIESCHE-STARNECKER^c

^a AG DigiHealth, Neu-Ulm University of Applied Sciences, Neu-Ulm, GERMANY

^b Technical University of Darmstadt, Darmstadt, GERMANY

^c Institute for Pathology, Technical University of Munich, Munich, GERMANY

^d IGHS, University of California, San Francisco, San Francisco, CA, USA

^e IBE, Ludwig Maximilian University of Munich, Munich, GERMANY

^f School of Medicine, Ulm University, Ulm, GERMANY

Keywords. Glioblastoma, Neural Networks, Expert Systems

1. Introduction

Glioblastoma multiforme (GBM), the most common type of malignant brain cancer in adults, has a poor prognosis with a median survival of around 15 months despite aggressive treatment.[1] GBMs are very heterogeneous tumors that differ in morphological and genetic aspects. Due to the genetic diversity, subtypes are described in the literature, that differ in their prognosis (a classical, mesenchymal, proneural, and occasionally neural subtypes).[2] Modern machine-based methods have a large potential for the automated analysis of medical images and are successfully used in neuropathology. [3] This study investigates if a neural network can automatically measure the heterogeneity of GBMs and can support prognosis estimates in a clinical setting.

2. Method

57 tissue samples of glioblastoma (Isocitrate dehydrogenase (IDH)-wild-type) have been fixed in 10% formalin and embedded in paraffin. 2 µm thick slices were stained with hematoxylin and eosin. 35 samples from brains from autopsies without pathological abnormalities have been prepared as a reference.

Slides were scanned using an Aperio AT2 scanner. Next, tumor tissue was digitally labeled and separated from the surrounding brain tissue. The first step will be to train a neural network for the delamination of GBM tissue from healthy tissue. As a next step, an additional neuronal network will be trained to identify the degree of

¹ Corresponding Author, Michael Oertl, E-mail: michael.oertl@hs-neu-ulm.de

morphological heterogeneity of glioblastoma. This data will be backed up by immunohistochemistry and presented as a color-coded image. A Convolutional Neural Network (CNN) will be developed and trained because it has already been used successfully on classification tasks for image analysis.[4]

3. Results

92 images have been investigated, which have an average size of 6 GB and overall 552 GB of data that needed to be processed. Input for the neuronal network is always a small area of the regarded image. The network shall classify this area. After every step, the area is shifted by a stride. The shifting enables us to keep the network smaller as we train on smaller data samples, and with this technique, we generate a vast amount of training data. For example, if we have an image with 70.000x40.000 pixels and the area we want to classify has the size of 200x200 pixels, and we shift the area by 100 pixels in each step, we have generated 278.901 training samples from one image. The output is a probability distribution over the possible classes. Firstly, these classes are GBM and non-GBM, and in the second step, the classes will be over the heterogeneity of the tumor.

4. Discussion

The authors suppose that a higher heterogeneity in the tumor tissue is related to worse survival chances of the individual. The application of neuronal networks for the neuropathology is still at its beginning. Nonetheless, automatic measures of the histological tumor heterogeneity could lead to more precise predictions and therefore, also to individual risk-adapted therapies over the long term.

5. Conclusion

We presented a new method for GBM detection and classification. This method enables a faster and automated tumor detection, as well as classification on the tumor to enhance individual prognoses.

References

- [1] Omuro A, DeAngelis LM. Glioblastoma and Other Malignant Gliomas. *JAMA* 2013;310:1842.
- [2] doi:10.1001/jama.2013.280319.
- [3] Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 2010;17:98–110. doi:10.1016/J.CCR.2009.12.020.
- [4] Xu Y, Jia Z, Ai Y, Zhang F, Lai M, Chang EI-C. Deep convolutional activation features for large scale Brain Tumor histopathology image classification and segmentation. 2015 IEEE Int. Conf. Acoust. Speech Signal Process., IEEE; 2015, p. 947–51. doi:10.1109/ICASSP.2015.7178109.
- [5] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks 2012:1097–105.

Automated Nerve Fibres Identification and Morphometry Analysis with Neural Network Based Tool in MATLAB

Michał KOPKA^a, Wiktor PASKAL^{a1}, Adriana M. PASKAL^a, Piotr PIETRUSKI^b, Ryszard KOPKA^c and Paweł K. WŁODARSKI^a

^a*Department of Methodology, Laboratory of Centre for Preclinical Research, Medical University of Warsaw, Warsaw, Poland*

^b*Timeless Plastic Surgery Clinic, gen. Romana Abrahama 18/322, 03-982 Warsaw, Poland*

^c*Department of Electrical Control and Computer Engineering Opole University of Technology, Opole, Poland*

Abstract. Analyses of nerve histology are core assays in basic and applied research and even in clinical setting. Detailed report on nerve morphology may unbiasedly indicate the current state of a peripheral nerve. Manual method requires trained technician and is a time-consuming procedure. Available plugins to well known image processors are limited in use and data outcomes. Thus, the aim of the study was to create a tool for fast and repeatable analysis of a nerve section image. As a results we get very high precision of analysis in shorter time.

Keywords. Axon evaluation, Neural network

1. Introduction

Preclinical research of neuroregeneration requires assessment of nerve fibers histology. The most reliable results contain histomorphometry parameters of axons and myelin sheaths. Manual methods are prone to human error and are time consuming. Computer-based techniques may improve the above and deliver the accuracy comparable with a experienced researcher. The study aimed to prepare a neural-network MATLAB script, with a novel approach to contrast-based method used for whole nerve section analysis.

2. Materials and methods

Semithin sections of a rat sciatic nerve stained with toluidine blue were a input data. They were scanned at 40x magnifications. Firstly, the analysis includes grey-scale conversion and contrast improving modifications e.g. CLAHE. Unlike other described methods, which analyze objects after binarization, we worked with the grey-scale image. Objects found with regionprops function were then filtered. Our protocol offers manual verification and removal of false-positive objects. We also applied Neural Net Pattern

¹ Corresponding Author E-mail: wikt.paskal@gmail.com

Recognition from Matlab for filtering results. Then, a rectangle enclosing each border of an object was defined. From the center to the edge of the acquired area line segment was lead and histogram on this length was set. The border of each axon was determined in the point of highest contrast between neighbour pixels. The borders of myelin sheaths were defined likewise. This operation was repeated 360 times, around the analyzed nerve cell (radial histogram scanning and thresholding). The method resembles manual analysis since local contrast is used to mark the axon.

Finally, the main morphometric parameters may be determined (radius, G ration, thickness, area etc.). Results were compared with manual measurements of 28 random nerve images, estimated by 3 researchers in ImageJ. Comparison to manual data was shown as ratios: script: manual.

3. Results

After optimization of brightness and contrast parameters, axon area overlay at level 1,001 and 1,009 of the myelin sheaths was achieved. Paired T-test showed no significant differences between these two methods (p -value = 0,321). Use of Neural Net enables to reach 93,6% of accuracy of axons recognition vs 3 researchers. A run time of the script is about 12 times shorter than the manual method.

4. Discussion

Other similar methods consist of binarization, which doesn't allow to define borders of the nerves precisely. The script presented shows a new approach to this case. It is based on reconstructing the way of evaluation done by the scientist. The presented script shows a new approach to this case. Our idea is based on reconstructing the way of evaluation done by the scientist. Owing to this innovative idea, the higher level of accuracy of computerized evaluation is possible. The most depending factor is optimising the differentiation between correctly and incorrectly marked borders. Further researches focus on increasing the efficiency of this protocol. In conclusion, the presented script performs an accurate analysis of nerve sections on a grey-scaled image. It overcomes bias of a binarization. The method decreases time of analysis and remains repeatable.

References

- [1] Moiseev, D., et al. (2018). "Morphometric analysis of peripheral myelinated nerve fibers through deep learning." *J Peripher Nerv Syst*. Isaacs, J., et al. (2014). "Modification of commercially available image analysis software for semi-automated qualitative analysis of axon regeneration and myelination in the rat sciatic nerve." *J Neurosci Methods* **233**: 45-49.
- [2] McCreedy, D. A., et al. (2016). "Semi-automated counting of axon regeneration in poly(lactide co-glycolide) spinal cord bridges." *J Neurosci Methods* **263**: 15-22
- [3] McCreedy, D. A., et al. (2016). "Semi-automated counting of axon regeneration in poly(lactide co-glycolide) spinal cord bridges." *J Neurosci Methods* **263**: 15-22
- [4] Patel, A., et al. (2018). "AxonTracer: a novel ImageJ plugin for automated quantification of axon regeneration in spinal cord tissue." *BMC Neurosci* **19**(1): 8.
- [5] Zaimi, A., et al. (2016). "AxonSeg: Open Source Software for Axon and Myelin Segmentation and Morphometric Analysis." *Front Neuroinform* **10**: 37.

Automatic Detection of Vital Signs in Clinical Notes of the Outpatient Settings

Martín DIAZ MAFFINI^{1a} and Fernanda AGUIRRE OJEA^a and Matías MANZOTTI^a

^a*Health Informatics Office, Hospital Alemán, Argentina*

Abstract. The determination of vital signs is a fundamental aspect of patient care. Electronic health records have a structured format for their registration. It is known that the frequency with which this data is recollected is not representative of reality. To complement the missing data we have created a tool that extract the information regarding blood pressure, heart rate, respiratory rate, height, weight and pain level recorded in free text in the clinical notes of outpatients.

Keywords. Natural Language Processing, Vital Signs, Electronic Health Records

1. Introduction

The determination of vital signs is a fundamental part of patient care in order to control, maintain or modify the therapeutic measures carried out. Most electronic health records have a structured format for the registration of this information, however, the frequency with which this data is recorded is not representative of reality [1]. Electronic health records capture data in two ways: in a structured format and in free text format. For the use of clinical data for secondary purposes the data must be structured. However, the exclusive approach of structured data for the entry of clinical data can result in the loss of significant information that is usually entered in free text format. [2,3] To complement the data of vital signs missing from the clinical notes we have created a tool that extracts this information regarding blood pressure, heart rate, respiratory rate, body temperature, height, weight and pain level recorded in free text of the clinical notes of outpatients.

2. Methods

The system extracts mentioned vital signs from the text of the clinical notes. These texts typically include ad hoc abbreviations, common words and spelling errors. In order to create the extraction system, a set of clinical notes was manually annotated by experts and analyzed to infer patterns. Such patterns were coded as regular expressions, each with a degree of specificity. The system attempts to detect vital signs by first applying more specific rules and then testing more generic rules along with additional validity checks to exclude false positives. The application of more sophisticated machine learning techniques was evaluated but rejected due to the requirement of both high precision and false positive detection in this specific problem. The results of these patterns and rules

¹ Correspondent author: martindiaz@hospitalaleman.com

in the clinical notes were evaluated by experts, who tried to find opposite examples and then refine them in an iterative way to increase accuracy and recall. The system was implemented within the UIMA (Unstructured Information Management Architecture)[4]. The analysis of each clinical note is comprised by several phases: first a phase where grammatical sentences are divided, followed by a tokenization phase which aims to detect running words and, finally, a vital signs detection phase. The detectors of each vital sign are implemented as independent components that the tool calls during the start of the process, so the tool is scalable and should be able to access new future components when available.

3. Results

For the phase of manual registration of vital signs in the clinical records, 3960 representative clinical notes of the different specialties that record information in search of the vital signs were evaluated. We identified in the free text: 2 records of pain level (PL), 101 heart rate (HR), 22 respiratory rate (RR), 35 body weight (BW), 225 blood pressure (BP), 7 of height (HT) and 53 of body temperature (BT).

Different configurations of the extraction tool were tested until obtaining a specificity and sensitivity of 100% against the training set.

The tool was implemented in such a way that it would be executed after 5 seconds of inactivity of the medical user in the outpatient electronic health record. The registration rate of structured vital signs data was measured one year prior to implementation and measured again after implementing the tool. The results were as follows: HR 448 occurrences, RR 122, BW 6935, BP 4156, HT 2610, BT 82, PL 15 and HR 4444, RR 1084, BW 8289, BP 9820, HT 3786, BT 1472, PL 1230 respectively. Increases were registered in the PL record 82 times more, BT 18 times more, HR 10 times more, RR 9 times more, BP 2.5 times more, HT 1.5 times more and BW 1.2 times more.

4. Conclusions

Vital signs registration in clinical notes is presented in patterns which can be determined and, therefore, rules can be developed to achieve their extraction by using tools that help achieve greater completeness of the records. The use of these tools increases the structured registration of these data, enabling their subsequent secondary use in clinical decision making.

5. References

- [1] Patterson OV, Jones M, Yao Y, Viernes B, Alba PR, Iwashyna TJ, DuVall SL. Extraction of Vital Signs from Clinical Notes. *Medinfo* 2015: 1035.
- [2] Barrows RC, Busuioc M, Friedman C. Limited parsing of notational text visit notes: Ad-hoc vs. NLP approaches. *Proc AMIA Symp.* 2000:51–5.
- [3] Hyun S, Bakken S, Friedman C, Johnson S. Natural language processing challenges in HIV/AIDS clinic notes. *AMIA Annu Symp Proc.* 2003:872.
- [4] <http://uima.apache.org/>. Accessed 11/20/2018

Best Practices for Designing Discrete Choice Experiments and the Use for Older Adults with Cognitive Impairment

Thomas ENGELSMA^{a, b, 1}, Monique W. M. JASPERS^{a, b}, Anne M. TURNER^{c, d}

^a Center for Human Factors Engineering of Health Information Technology,
^b Department of Medical Informatics, University of Amsterdam, Amsterdam, The Netherlands;

^c Department of Health Services, School of Public Health, University of Washington, Seattle, WA; ^d Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA.

Abstract. *Objective:* the aim of this scoping review is to 1) identify guidelines, frameworks, issues, and recommendations regarding the development of a discrete choice experiment (DCE), 2) look at potential design considerations regarding the development of a DCE for older adults with dementia, and 3) summarize knowledge about the current inclusion of and experiences with older adults with cognitive impairment in DCE studies. *Methods:* Two literature searches divided into two topics were performed in PubMed, CINAHL, and PsycINFO. *Results:* Topic 1 found 19 articles regarding frameworks, guidelines and design considerations for DCEs in healthcare. Topic 2 found eight articles regarding the effect of cognitive impairment on DCE results and/or discussing the inclusion of cognitive impaired older adults in these DCE studies. *Conclusion:* There is minimal literature available regarding the effect of cognitive impairment on DCE results. In future studies we will explore if DCEs can be used to involve older adults with dementia in decisions regarding transitions in care.

Keywords. Discrete choice experiment, design, older adults, cognitive impairment

1. Introduction

There are a growing number of older adults worldwide. Older adults are at increased risk of developing cognitive problems and dementia. Management of dementia often requires making a variety of important care-related decisions. These decisions often occur without the input of the older adult suffering from dementia or another severe illness [1]. It has been shown that including older adults with dementia in decision-making can lead to an improved well-being and reduce dementia-related symptoms [2]. This literature review was conducted as preliminary work to inform the development of a DCE-based tool for older adults with dementia to support their participation in the decision-making process.

¹ Corresponding Author email: t.engelsma@amsterdamumc.nl

2. Methods

We conducted two literature searches regarding DCEs. The first topic (Topic 1) focused on lessons learned from applications of DCEs in healthcare. The second topic (Topic 2) focused on DCE studies including older adults with cognitive impairments. PubMed, CINAHL, and PsycINFO were searched for both topics. Studies had to be published within 2009-2019 and written in English. For inclusion in Topic 1 articles had to discuss the design process of a healthcare-related DCE. For Topic 2, articles were included that 1) describe DCE studies targeting older adults (65 and over) and 2) discussed the effect of cognitive impairment on DCE results and/or mentioned cognitive impairment in their inclusion process.

3. Results

A variety of frameworks, guidelines, and design considerations were identified in Topic 1 that might be useful when developing DCEs for older adults with dementia, such as using visuals. Minor literature regarding the effect of cognitive impairment on the results of a DCE was found in Topic 2. The few studies involving the use of DCE in older adults with cognitive impairment, suggest that older adults with mild cognitive impairment are equally capable of giving meaningful responses compared to those who are cognitively intact. However, these studies had a small sample size. A small number of studies discussed how they went about including older adults with cognitive impairment in their DCE. Most of these studies involved individuals with only mild cognitive impairment based on a cognitive assessment, or based inclusion on the opinion of professionals or ability to provide informed consent.

4. Discussion and Conclusion

This scoping review summarized insights on DCEs in healthcare with the aim of informing the design of a DCE for older adults with dementia. We found that in healthcare related context a minority of researchers included older adults with cognitive impairments in their DCE. Older adults with more severe levels of cognitive impairment are either completely excluded from participation in DCEs, or proxies (usually caregivers) participate in the DCE. However, studies suggest that caregivers' preferences regarding healthcare often differ from the recipient's preferences [3]. Therefore, DCE data may not be reliable when using proxies. In future studies we will explore if DCEs can be used to involve older adults with dementia in decisions regarding transitions in care. The final aim is to develop a DCE- based digital decision-making tool for older adults with dementia and their caregivers.

References

- [1] Groen van de Ven L, et al. Recognizing decision needs: first step for collaborative deliberation in dementia care networks. *Patient Educ Couns.* 2017;100(7):1329-1337.
- [2] Menne HL, Tucke SS, Whitlatch CJ, Feinberg LF. Decision-making involvement scale for individuals with dementia and family caregivers. *Am J Alzheimers Dis Other Dement.* 2008;23(1):23-9.
- [3] Smebye KL, Kirkevold M, Engedal K. How do persons with dementia participate in decision making related to health and daily care? a multi-case study. *BMC Health Serv Res.* 2012;12:241.

Biostatistics Disruptive Acculturation Through Serious Gaming: A New Hope

Pierre GILLOIS ^{a,1}, Lionel DI MARCO ^a, Josep M. NICOLÀS ^b,

Alexandre MOREAU-GAUDRY ^a, Anne EGO ^a, Sandra DAVID-TCHOUDA ^a,

Donald K MARTIN ^c, Jean-Luc BOSSON ^a

a CNRS / TIMC-IMAG Lab., UMR 5525 (Themas), Grenoble-Alpes University, France

b IDIBAPS, University of Barcelona, Barcelona, Spain

c CNRS / TIMC-IMAG Lab., UMR 5525 (SyNaBi), Grenoble-Alpes University, France

Abstract. Biostatistics is one of the transversal subjects that all future doctors must acquire and master. Nonetheless, it is a subject that has the reputation of being difficult, which has not been able to be corrected even with the application of new pedagogical methods such as blended learning. We address this problem with our acculturative and disruptive approach in the form of a serious game scenario in clinical research that integrates biostatistics with our R4Web adapted tools. Our approach was launched in 2008 for the second year of medical school. Here we describe this LOE scenario for serious game including the biostatistics disruptive acculturation task and present its new international version.

Keywords. “Schools, Medical”, Bio-statistics, Serious games, “Education, Public Health Professional”, “Games, Experimental”, “Data Interpretation, Statistical”

1. Introduction: Learning context

Some subjects are part of these transversal competencies that all modern health professionals must master. Biostatistics is one of those competencies, that provides the necessary foundation for understanding and analysis in all health fields whether through the critical appraisal of articles (CAA). Major efforts have been made to integrate this transversal subject into the initial medical training in the form of a revised pedagogy using e-learning, blended learning, etc. [1,2].

Thus, it behoves health educationalists to develop pedagogical strategies that overcome the barrier of disinterest and misunderstanding on the part of students, who often consider it too mathematical, useless for health and medical practice and do not realize its underlying and real importance for their professional lives. To overcome this challenge, in 2008 a serious game simulating a clinical research process has been developed in Grenoble. The serious game scenario is called “LOE” (Laboratorium Of Epidemiology) is hosted on the multilingual health Grenoble Serious Games platform: Professional Health Fantasy (<https://prohf.org>).

¹ Corresponding Author,: pierre.gillois@univ-grenoble-alpes.fr.

2. Materials: Serious game scenario description

Provided by the platform, interaction activities of the scenario are integrated in the form of online forms, emails, integrated pseudo websites such as the Medical Information Department for the data request, or the International Conference for the final step. Collaborative activities for students, groups or tutors are also supported by the platform. This platform therefore centralizes all the information and the digital tools required to simulate clinical research in all these dimensions.

This LOE serious game scenario immerses students in a public health physicians' situation, since it involves investigating the incidence of a disease: venous thromboembolic disease (VTE) and that includes all the steps required to undertake a real clinical research program. The main objectives for students of this situation are to (i) design the protocol and implement an epidemiologic study, (ii) prepare and submit a scientific paper including the biostatistical analysis task (with the adapted R4Web tool: <https://r4web.univ-grenoble-alpes.fr>), to describe the protocol and epidemiological study, for presentation at a simulated conference; (iii) finally propose a decision-making tool to optimize the VTE diagnostic process.

3. Practice Analysis: User experience

By discovering the game, the learner will wander through the various virtual tours of the group's hospitals to discover the units, services, resources and tools at his disposal for the mission. The IT environment is designed to stage different interactions with scenario characters in the game, such as patients in the hospital, PPC experts, members of the conference scientific committee, etc. This in order to help students to take on different challenges and acquire skills related to the scenario, by engaging into the game. Over time of the game, the students are judged on the quality of their production: protocol, article and keynote. The role of the teachers as tutors is to only guide the learners during the classroom sessions. The success of this serious game depends on the need for involvement and commitment of both learners and tutors. The challenge is that learners and tutors must leave their comfort zone and must go beyond with their computer.

Finally, informal feedback from our clinical colleagues is that the tools that utilise real research projects provide immediate and positive implications for the young doctors in-training. All the feedback reinforces the benefits of this new approach, so that biostatistics is no longer an unavoidable cliff but just a fun step as part of an overall clinical health research process. This active approach to problem-based learning is necessarily more "disruptive" for students and a challenge for tutors.

4. References

- [1] A.J. Rock, W.L. Coventry, M.I. Morgan, and N.M. Loi, Teaching Research Methods and Statistics in eLearning Environments: Pedagogy, Practical Examples, and Possible Futures, *Front Psychol.* 7 (2016). doi:10.3389/fpsyg.2016.00339.
- [2] P. Gillois, J.L. Bosson, C. Genty, J.-P. Vuillez, and J.-P. Romanet, The impacts of blended learning design in first year medical studies, *Stud Health Technol Inform.* 210 (2015) 607–611.

Challenges of In-House Development and Implementation of a CPOE for Oncology

S. de MATOS LIMA^a, C L GIMENEZ^a, D. LUNA^a

^a *Department of Health Informatics, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina*

Abstract. Hospital Italiano de Buenos Aires (HIBA) is an academic tertiary care hospital highly specialized that has started the process of informatization of chemotherapy protocols. The objective is to describe the development of a computerized physician order entry (CPOE) oriented to the oncology adult patient and the members of the healthcare team that works with him (physicians, pharmacists, nurses and administrative staff) to improve the process and prevent errors at a critical point in the patient's health care: during prescription, preparation and / or administration. The development of this system consisted of several stages: inquiries about the usual work and perception of needs of the potential users; user-centered design; interoperability with the electronic health record (EHR) and development of a final prototype

Keywords. CPOE, oncology, chemotherapy

1. Introduction

The oncological patient represents a healthcare challenge [1] Patient safety is of vital importance and this is where information systems can reduce the rate of errors and adverse events. [2] The prescription cycle in oncology turns out to be complex due to the coexistence of multiple structured schemes and the high toxicity of drugs.[3] The benefits of implementing a CPOE for indication of medications have been described, from the reduction of errors to integration of clinical decision support systems. [4] and some have shown their long-term benefit [3, 4] One of the challenges in oncology is to prevent errors that may occur during the different phases of the prescription cycle[5] Some appears to be associated with more significant adverse outcomes, with approximately one third of all errors identified as serious or life-threatening [5]. Methods A descriptive observational study was carried. The study is a mixed design, with a qualitative and quantitative approach.

2. Results

In October 2012 we started to develop and implement a CPOE integrated into the hospital's EHR, planning to incorporate pediatric patients at a later stage [7] For the start a surveyed and meetings were held between the project team. In a first stage in the standardization of order sets we start working in spreadsheet format for a year. The team was made up of one health informatics specialist, one process analyst and two developers. Six people participated in the preparation committee of order sets: one clinical oncologist, one hematologist, one gynecologist, head of oncology pharmacy, and oncology nursing supervisors. This phase lasted 7 months. By July 2013, the 50 most used protocols in the institution were completed, agreed between the prescribing services, oncology pharmacy

and nursing. In October 2013 there were 100 standardized protocols. The second stage began with the development of the RCM to structure the order sets, add agendas in the ADT for the sectors that treat cancer patients, connect the prescription module with the pharmacy administration system (PAS), incorporate the preparation processes of the oncology pharmacy into the PAS, and modify the EHR to allow the prescription of oncological protocols. Finally, the design was completed and it was implemented in November 2014.

3. Discussion

Protocols were added up until reaching almost all of those available for adults. The implementation was followed by a process of continuous improvements and system adjustments that allowed a superior user experience and avoid frequent errors that were detected. As a strength, the in-house development that allowed the system to adapt to the previous workflow and fully interoperate with the HER.

4. Conclusions

We have described the process of implementing a CPOE of chemotherapy in a high complexity hospital with its own development team and how it was arrived from a paper model to another that interoperates with 4 simultaneous systems in a complex circuit in which physicians, pharmacists, nurses and administrators interact in different stages, which are often critical for the patient.

References

- [1] Dow LA, Matsuyama RK, Ramakrishnan V, et al. Paradoxes in advance care planning: the complex relationship of oncology patients, their physicians, and advance medical directives. *J Clin Oncol*. 2009;28(2):299-304.
- [2] Margaret H. Reckmann, Johanna I. Westbrook, Yvonne Koh, Connie Lo, Richard O. Day; Does Computerized Provider Order Entry Reduce Prescribing Errors for Hospital Inpatients? A Systematic Review, *JAMIA*, Volume 16, Issue 5, 1 Sept 2009, Pages 613–623, <https://doi.org/10.1197/jamia.M3050>
- [3] Carrington, Christine, et al. «The Clinical Oncological Society of Australia (COSA) Guidelines for the Safe Prescribing, Dispensing and Administration of Cancer Chemotherapy». *Asia-Pacific Journal of Clinical Oncology*, vol. 6, n.º 3, septiembre de 2010, pp. 220-37. Crossref, doi:10.1111/j.1743-7563.2010.01321.x.
- [4] Greenberg, Anna, et al. «Cancer Care Ontario's Computerized Physician Order Entry System: A Province-Wide Patient Safety Innovation». *Healthcare Quarterly (Toronto, Ont.)*, vol, 2006, pp. 108-13.
- [5] Elsaid KA, Garguilo S, Collins CM. Chemotherapy e-prescribing: opportunities and challenges. *Integr Pharm Res Pract*. 2015;4:39-48. Published 2015 May 21. doi:10.2147/IPRP.S84232
- [6] Daniel L, Leonardo G, Bernaldo G, Fernán DQ, Adrián G. Desarrollo e implementación de un Sistemas de Prescripción Electrónica. 2001;(June 2014).
- [7] Kim GR, Chen AR, Arceci RJ, et al. Error Reduction in Pediatric Chemotherapy Computerized Order Entry and Failure Modes and Effects Analysis. *Arch Pediatr Adolesc Med*. 2006;160(5):495–498. doi:10.1001/archpedi.160.5.495

Classification of Biceps Brachii Muscle Fatigue Condition Using Phase Space Network Features

Navaneethakrishna MAKARAM^{a1} and Ramakrishnan SWAMINATHAN^a

^a*Non-Invasive Imaging and Diagnostics laboratory, Department of Applied Mechanics
Indian Institute of Technology Madras, Chennai, India - 600036*

Abstract. In this, study, an attempt is made to differentiate muscle nonfatigue and fatigue condition using signal complexity metrics derived from phase space network features. A total of 55 healthy adult volunteers performed dynamic contraction of the biceps brachii muscle. The first and last curl are segmented and are considered as nonfatigue and fatigue condition respectively. A weighted phase space network is constructed and reduced to a binary network based on various radii. The mean and median degree centrality features are extracted from these networks and are used for classification. The results of the classification indicate that these features are capable of differentiating nonfatigue and fatigue condition with 91% accuracy. This method of analysis can be extended to applications such as diagnosis of neuromuscular disorder where fatigue is a symptom.

Keywords. Muscle fatigue, Phase space networks, Multilayer perceptron

1. Introduction

Neuromuscular fatigue can be defined as a task-induced reduction of the force-generating capacity of a muscle. Surface Electromyography (sEMG) signal reflects the occurrence and progression of muscle fatigue. Conventional signal processing techniques fail to quantify muscle fatigue. In this study, nonlinear features are extracted using phase space networks to explore the dynamics of the complex biological system.

2. Methods

The study involves subjects performing biceps curl exercise with a six kilogram force dumbbell until fatigue. During this period sEMG signals are recorded from the biceps brachii muscle in accordance with the SENIAM standards. A data acquisition rate of 10ksp/s and a gain of 1000 is used. These signals are preprocess with a 10-400 Hz band pass and a 50 Hz notch filter. The first and last curl are extracted based on the methodology mentioned in [1]. The phase space is constructed based on Takens embedding procedure with a dimension of 3 using false 1 neighborhood procedure.

¹ corresponding author: email kmakaram@yahoo.com

The weighted phase space network is constructed based on the following criterion [2].

$$W_{ij} = |X_i - X_j| \tag{1}$$

Where, W_{ij} is the adjacency matrix, and, X_i and X_j correspond to the instance in the phase space and represent a node. The binary network is constructed by using the following condition:

$$W_{ij} \leq r \tag{2}$$

Here, the threshold value r ranges from 0.05 to 0.5 in steps of 0.05 times the standard deviation of the signal. This binary network degree distribution is computed. The mean and median degree centrality measures are extracted. The classification is performed using a multilayer perceptron (MLP) algorithm and a 10 fold cross validation scheme.

3. Results and Discussion

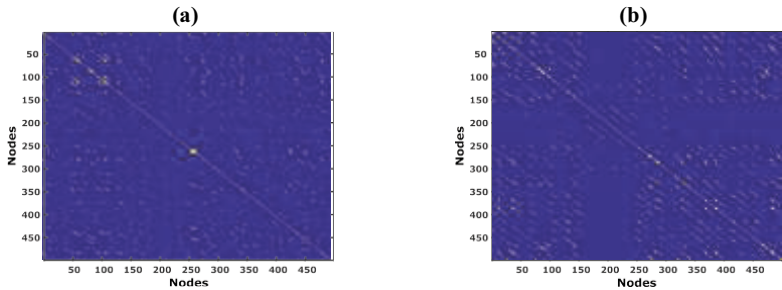


Figure 1. Representative adjacency matrix obtained with an r value of 0.25SD (a) nonfatigue and (b) fatigue

In figure 1, the adjacency matrix obtained from a subject in nonfatigue and fatigue condition is shown. It is seen that the fatigue condition has a greater number of connections. The centrality features indicate the presence of a greater number of recurrent (similar) patterns in fatigue. The classification performance of the extracted features are shown in table 1. It is seen that mean degree centrality and median degree centrality separately perform well with an average accuracy greater than 85%. When both these features are used in conjunction an accuracy of 91% is obtained. It is to be noted that Median centrality feature alone gives similar performance where in a single subject is misclassified in this case leading to the disparity in the performance.

Table 1 Classification performance of MLP for various features

Features		Sensitivity	Specificity
Mean Centrality	85.7	85.7	85.7
Median Centrality	89.3	87.9	90.7
Both	91.0	89.7	92.6

References

[1] P. A. Karthick, D. M. Ghosh, and S. Ramakrishnan, Surface electromyography based muscle fatigue detection using high-resolution time-frequency methods and machine learning algorithms, *Computer Methods Programs Biomedicine* **154** (2018), 45–56.

[2] B. Henry, N. Lovell, F. Camacho, Nonlinear dynamics time series analysis, M. Akay Nonlinear Biomedical Signal Processing: Dynamic Analysis and Modeling, IEEE Press, New York, 2001.

Classification of First-Episode Schizophrenia Using Wavelet Imaging Features

Kateřina MARŠÁLOVÁ^{a,b}, Daniel SCHWARZ^{a,b} and Ivo PROVAZNÍK^c

^a Masaryk University, Faculty of Medicine, Kamenice 5, 62500 Brno, Czech Republic

^b Institute of Biostatistics and Analyses, Ltd., Czech Republic

^c Brno University of Technology, Technická 3058/10, 61600 Brno, Czech Republic

Abstract. This work explores the design and implementation of an algorithm for the classification of magnetic resonance imaging data for computer-aided diagnosis of schizophrenia. Features for classification were first extracted using two morphometric methods: voxel-based morphometry (VBM) and deformation-based morphometry (DBM). These features were then transformed into a wavelet domain using the discrete wavelet transform with various numbers of decomposition levels. The number of features was then reduced by thresholding and subsequent selection by: Fisher's Discrimination Ratio (FDR), Bhattacharyya Distance, and Variances (Var.). A Support Vector Machine with a linear kernel was used for classification. The evaluation strategy was based on leave-one-out cross-validation.

Keywords. Machine learning, neuroimaging, schizophrenia, support vector machines

1. Introduction and methods

Machine learning offers promising methods for computer-aided diagnosis of neuropsychiatric disorders. The use of these computing methods is highly desirable, as they have the potential to create an objective process of early diagnosis and prognosis. This research explores the design of the classification algorithm and the application of wavelet transformation to image data. Our proposed algorithm follows the ideas in [1]. Compared to the original algorithm, the main advantage of our version lies in the ability to carry out more experiments with different decomposition levels by wavelet transform.

The algorithm consists of five main steps: (i) Two types of preprocessing were used to extract features: VBM (gray matter densities) and DBM (local volume changes), (ii) images were decomposed by wavelet transformation into another domain. It was necessary to specify two parameters: decomposition levels and mother wavelet (sym5) [2], [3], (iii) the number of features was reduced by the experimentally determined thresholding of absolute values smaller than 0.05, (iv) the number of features was reduced by selection. Three different metrics were used for this: FDR, Bhattacharyya distance, and Var., (v) the classification into two groups (FES/NC) was carried out. The algorithm was tested on a dataset provided by The University Hospital Brno (104 T1-weighted MR images of subjects, of which 52 subjects were patients with first episode

of schizophrenia (FES) and 52 were healthy volunteers (HC)). Classification performance was evaluated using leave-one-out cross-validation. The best results of experiments with differently selected parameters are shown in Table 1.

Table 1. Classification accuracy for distinguishing between FES and HC.

Morphometric method	Metrics	Decomposition level	Number of features	Accuracy [%]	Sensitivity [%]	Specificity [%]
DBM	FDR	4	1000	56.73	55.77	57.69
			5000	69.23	67.31	71.15
		5	1000	58.65	57.69	59.62
			5000	72.12	67.31	76.92
	Var	4	1000	57.69	55.77	59.62
			5000	69.23	67.31	71.15
		5	1000	59.62	57.69	61.54
			5000	71.15	67.31	75.00
VBM	FDR	4	1000	73.08	71.15	75.00
			5000	67.31	65.38	69.23
		5	1000	71.15	69.23	73.08
			5000	67.31	63.46	71.15
	Var	4	1000	73.08	71.15	75.00
			5000	69.23	67.31	71.15
		5	1000	71.15	69.23	73.08
			5000	68.27	65.38	71.15

2. Discussion and conclusion

The strongest results were obtained using DBM, where the classifier achieved the best performance for five decomposition levels selected using FDR and Var for 5000 features. The first case achieved 72.12% accuracy, 67.31% sensitivity and 76.92% specificity, and the second case attained 71.15% accuracy, 67.31% sensitivity and 75.00% specificity. Using VBM with four decomposition levels for 1000 features and Var. and FDR metrics, 73.08% accuracy, 71.15% sensitivity, and 75.00% specificity were achieved.

The results from this analysis are comparable to those of similar works [4], [5], [6]. Despite the classification of FES and HC, the resulting accuracy is still too low to propose the algorithm be used in clinical practice as an objective method for the diagnosis of this devastating neuropsychiatric disorder. This work was supported by the project No. AZV 17-33136A (Ministry of Health, CZ).

References

[1] P. Dluhoš, *Multiresolution feature selection for recognition in magnetic resonance brain images*. Brno, 2013. Master thesis. Masaryk University, Faculty of Science, Department of Experimental Biology.

[2] N. K. Al-Qazzaz, et al., Selection of mother wavelet functions for multi-channel EEG signal analysis during a working memory task, *Sensors* **15** (2015), 29015–35.

[3] S. Kumari, Effect of Symlet filter order on denoising of still images, *Advanced Computing: An International Journal* **3** (2012), 137–143.

[4] D. Schwarz, et al., A deformable registration method for automated morphometry of MRI brain images in neuropsychiatric research, *IEEE Transactions on Medical Imaging* **26** (2007), 452–461.

[5] J.-L. Starck, et al., *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press, 2010.

[6] J. H. Yoon, et al., Automated classification of fMRI during cognitive control identifies more severely disorganized subjects with schizophrenia. *Schizophrenia Research* **135** (2013), 28–33.

Classification of Histologic Images Using a Single Staining: Experiments with Deep Learning on Deconvolved Images

Vincenzo DELLA MEA^{a,1} and David PILUTTI^a

^a*Dept. Of Mathematics, Computer Science and Physics, University of Udine, Italy*

Abstract. The automated analysis of digitized immunohistochemistry microscope slides is usually a challenging task, because markers should be analysed on the tumor area only. Tumor areas could be recognized on a different slide, stained with Haematoxylin-Eosin. The basic idea of the present poster is to evaluate how well deep learning methods perform on the single haematoxylin component of staining, with the prospective possibility of developing a classifier able to recognize tumor areas on IHC slides on their haematoxylin component only. In a preliminary experiment, single stain images obtained by H-E color deconvolution showed an accuracy of 0.808 and 0.812 for Hematoxilyn and Eosin components, respectively.

Keywords. Deep Learning, Digital slides, cancer

1. Introduction

The automated analysis of digitized immunohistochemistry (IHC) microscope slides is an actual and challenging problem, because markers should be analysed on the tumor area only. The IHC samples are stained with diaminobenzidine (DAB) to reveal a specific biological marker, and counterstained with Haematoxylin (H) to identify the cellular context. Tumor areas could be recognized on a different slide, stained with Haematoxylin Eosin (H-E). This normally raises the complexity of an automated analysis, introducing new tasks such as the alignment of the differently counterstained samples. In the last years, however, notable results have been obtained in tumor and metastases detection on H-E stained digital slides using deep learning techniques [1,2].

The basic idea of the present poster is to evaluate how well deep learning methods perform on the single stainings, with the prospective possibility of developing a classifier able to recognize tumor areas on IHC slides on their H component only. A preliminary experiment has been made on H-E stained images coming from the Bach Challenge [2].

2. Methods

The dataset for the classification tasks comes from the BACH Challenge [2], Part A. It consists of training (n=400) histologic images from breast cancer, acquired at 0.42µm

¹ Corresponding Author, Vincenzo Della Mea, Dept. Of Mathematics, Computer Science and Physics, University of Udine, via delle Scienze 206 – 33100 Udine Italy E-mail: vincenzo.dellamea@uniud.it.

and labelled with 4 categories, imagewise (normal, benign, in-situ carcinoma, invasive carcinoma). As a baseline, a model has been trained on the original, full color images, split into 1024x1024 tiles with a 128 pixels stride, using Fastai as programming framework [3]. Images were then transformed into their H and E components, using the Color Deconvolution plugin of ImageJ [4] with its standard H-E vectors.

The same model has been trained separately on the two staining components, after RGB conversion to make them compatible with the model. Accuracy evaluation has been carried out with the Bach Challenge rules [2].

3. Results

The baseline model was based on ResNet152, pretrained on Imagenet on 3 epochs on the last layer, plus 3 epochs for fine tuning on the whole network. The model achieved almost state-of-the-art accuracy (average of 0.845 on 4 runs, with a maximum of 0.88). The Haematoxylin model reached an average accuracy of 0.8075 on 4 runs (maximum: 0.84), while the eosin model reached 0.812 (maximum: 0.83).

4. Discussion

The models trained on single staining images obtained an accuracy lower than the full image one, although, if compared with the challenge results, they could still rank among the best 10 [2]. This result has been obtained without any specific optimization for these images, as the model architecture and hyperparameters of the baseline model have been fully replicated. Furthermore, no custom deconvolution vectors could have been computed, and this could have lowered the classification performance.

Further work includes running the experiment on images acquired in a controlled environment, to compute the deconvolution vectors. Furthermore, classification of deconvolved H-DAB images in their Haematoxylin component will be also considered to prove a cancer recognition method for IHC in alternative to image registration

5. Acknowledgements

The Titan Xp used for this research was donated by the NVIDIA Corporation. Further training has been conducted on GCP thanks to Google Research Credits program.

References

- [1] B. E. Bejnordi, M.Veta, J.P.van Diest P, et al, Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer, *JAMA* **318**(2017), 2199-2210.
- [2] G.Aresta, T.Araujo, S.Kwok, et al, BACH: Grand challenge on breast cancer histology images, *Med Image Anal.* **56** (2019),122-139.
- [3] Howard et al, FASTAI. URL: <https://docs.fast.ai/>
- [4] A.C. Ruifrok, D.A.Johnston, Quantification of histochemical staining by color deconvolution, *Anal. Quant. Cytol. Histol.* **23** (2001), 291-299.

Classification of Normal and Cardiomegaly Conditions in Chest Radiographs Using Cardio-Mediastinal Features

Satyavratn GOVINDARAJAN¹, Sukanta Kumar TULO and Ramakrishnan SWAMINATHAN

*Non-Invasive Imaging and Diagnostics Laboratory, Department of Applied Mechanics
Indian Institute of Technology Madras, Chennai, India - 600036*

Abstract. In this study, an attempt has been made to differentiate normal and cardiomegaly using cardio-mediastinal radiometric features and machine learning approaches. A total of 60 chest radiographs including normal and cardiomegaly subjects are considered from a public dataset. The images are preprocessed using edge aware contrast enhancement technique to improve the edge contrast of lung boundaries. The mediastinal, cardiac and thoracic widths and their radiometric indices are computed to characterize the morphological variations. The features are fed to three different classifiers for the differentiation of normal and cardiomegaly. Results show that the Linear discriminant analysis classifier is found to perform better with average values of recall 88.7%, precision 88.8%, and area under the curve 91.9%. Hence, the proposed computer aided diagnostic approach appears to be clinically significant to distinguish normal and cardiomegaly especially in remote and resource - poor settings.

Keywords. Cardiomegaly, Cardio-Mediastinal features, Machine learning

1. Introduction

Cardiomegaly is a medical condition that affects the cardiovascular system, specifically involving the enlargement of heart. This signature results from hosts of other diseases and strongly associated with congestive heart failure. Mediastinum region in the chest is considered as one of the substantial anatomical regions for the gross diagnosis of cardiovascular abnormalities. Early detection with mass screening approaches are crucial to prevent such condition, particularly in remote and resource-poor settings. Chest radiographs (CXRs) have been widely used for this purpose due to their non-invasiveness, and portability. Conventional image processing methods applied on CXRs fail to characterize the condition and differentiate them from normal subjects.

2. Methods

The frontal CXRs of 60 subjects (Normal- 30, Cardiomegaly-30) are obtained from Chest X-ray14 publicly available dataset. The images are in grayscale with a resolution

¹ Satyavratn Govindarajan, Non-Invasive Imaging and Diagnostics Laboratory, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, India - 600036; E-mail: s.vratan@gmail.com.

of 1024*1024. The images are subjected to edge aware contrast enhancement technique. It is used to enhance the contrast in smaller sub-regions of the image by retaining the edge information [1]. This method uses local Laplacian filtering based on Gaussian to distinguish edges from details. Sharpness measure is used to validate the edge contrast enhancement of lung boundaries.

Left mediastinal width and mediastinal width [2], cardiac width and thoracic width features are extracted, and their corresponding ratiometric indices such as Mediastinal width ratio, Cardiothoracic ratio [3], Mediastinal cardiac ratio (Mcr) and Mediastinal thoracic ratio are derived for all the subjects. All the features are provided to Linear Discriminant Analysis (LDA), Naïve Bayes (NB) and Support Vector Machine (SVM) classifiers. Ten-fold cross validation is performed and the performance measures are evaluated.

3. Results and Discussion

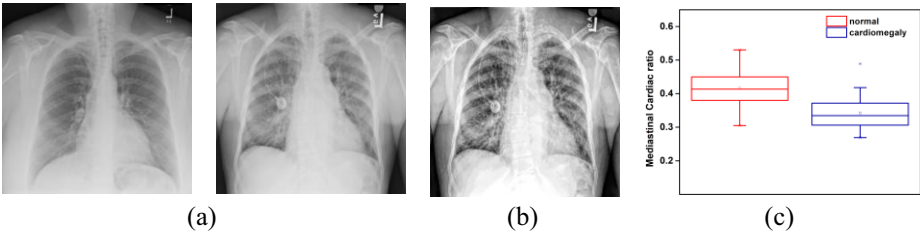


Figure 1. Representative images (a) normal and cardiomegaly (b) preprocessed (c) Mcr plot

From Figure 1, it is seen that the original images have poor edge contrast especially along lung boundaries at significant anatomical landmarks and the adopted preprocessing method is found to enhance the edge contrast locally. Significant distinction is observed between normal (0.3-0.55) and cardiomegaly (0.28-0.42) using Mcr plot.

The performance comparison of classifiers is shown in table 1. It is found that LDA performs better than NB and SVM with average values greater than 88% for recall, precision and F-measure and AUC of 91.9%. Large number of input images would be considered for future work.

Table 1. Performance of classifiers using all the features

Classifiers	Recall (%)	Precision (%)	F-measure (%)	AUC (%)
LDA	88.7	88.8	88.7	91.9
NB	82.3	82.4	82.4	88.4
SVM	85.5	85.5	85.5	86.0

References

[1] X.Y. Li, Y. Gu, S.M. Hu, and R.R. Martin, Mixed-domain edge-aware image manipulation, *IEEE Transactions on Image Processing* **22** (2013), 1915-1925.

[2] V. Lai, W.K. Tsang, W.C. Chan, and T.W. Yeung, Diagnostic accuracy of mediastinal width measurement on posteroanterior and anteroposterior chest radiographs in the depiction of acute nontraumatic thoracic aortic dissection, *Emergency Radiology* **19** (2012), 309-15.

[3] M.A. Hasan, S.L. Lee, D.H. Kim, and M.K. Lim, Automatic evaluation of cardiac hypertrophy using cardiothoracic area ratio in chest radiograph images, *Computer Methods and Programs in Biomedicine* **105** (2012), 95-108.

Classification of the Severity of Adverse Drugs Reactions

Raphaël CHAUVET^{1,2}, Cédric BOUSQUET¹, Agnès LILLO-LELOUET³, Ilan ZANA¹, Ilan BEN KIMOUN¹, Marie-Christine JAULENT¹

¹ Sorbonne Université, INSERM, Université Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, Paris, France.

² EISTI, International Engineering School, 95000 Cergy-Pontoise, France

³ Centre Régional de Pharmacovigilance, Hôpital Européen Georges-Pompidou, Assistance Publique – Hôpitaux de Paris, Paris, France.

Abstract. This poster presents a non-exhaustive study of machine learning classification algorithms on pharmacovigilance data. In this study, we have taken into account the patient's clinical data such as medical history, medications taken and their indications for prescriptions, and the observed side effects. From these elements we determine whether the patient case is considered serious or not. We show the performances of the different algorithms by their precision, recall and accuracy as well as their learning curves.

Keywords: Classification, Machine Learning, Pharmacovigilance, Decision support tool.

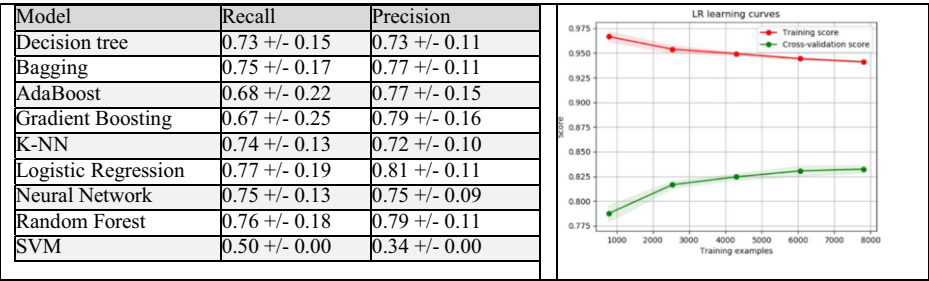
1. Introduction

Since 2017 in France, citizens can report directly to health authorities any adverse event related to health products. The objective of the presented work is to study to which extend machine learning algorithms could predict the seriousness of the situation from the description of the situation by patients. We used structured data recorded by health professionals from the verbatim of the patient to train machine learning algorithms. The main perspective is to apply the resulting prediction system to social media for pharmacovigilance monitoring. Although, signal detection in pharmacovigilance is historically based on statistical methods, more and more machine learning can make predictions for the detection of side effects [1]. Andrew M Wilson in [2] concluded that machine learning will not replace traditional pharmacovigilance technics but could collaborate to detect uncommon drug-related effects or reducing adverse drugs effects identification time.

2. Methods and Results

To carry out the study we considered a database that includes all the cases reported by patients from 2010 to 2019 resulting in a base of 13026 patient cases described by 50 variables from HEGP (European Hospital Georges Pompidou), Paris, France. Cases reported in this database are standardized by pharmacovigilant using MedDRA terms. 8699 cases are annotated as serious by health professionals. The aim of the study is to search the most efficient supervised learning algorithm that will enable the pharmacovigilant to prioritize serious cases. In a first step, all values of a variable are considered as a bag of word and translated in terms of presence/absence ("one-hot-encoding"). This operation allows to keep the inter-variable interaction for poly-medical or poly-pathological patients.

Before applying an algorithm, the dataset is divided in two, 75% of the database will serve as a training base and 25% will be used to test the learning performance. This process is repeated 5 times in order to have a general tendency of the performances of the different methods. Then we apply 9 different state-of-art clustering models to our dataset thanks to available libraries: *Decision Tree*, *Bagging Tree*, *Boosting Tree (AdaBoost)*, *Gradient Boosting Machine*, *K-Nearest Neighbours (K-NN)*, *Logistic Regression*, *Neural Network*, *Random Forest*, *Support Vector Machine (SVM)*. The next figure left shows the results in terms of precision (proportion of good prediction if the case is severe) and recall (percentage of serious cases correctly identified). The model with the best performance in terms of accuracy and recall is the logistic regression (next figure right). It does not over-fit the data and is very accurate (cf. LR learning curves next figure right).



3. Discussions and conclusion

It is important to note that the performance of machine learning methods depends on the input data that is used for learning. The same method on a different data set will give different results. Similarly, as new several observations are saved to the database, the performance of the same method will change over time. During this first study, it was decided to work initially with the default values of the algorithms because we did not have a priori knowledge to force the value of certain parameters. One possible improvement will be to choose optimal settings for all models and re-compare their performances later. For that we could use a search grid where all the parameters would be stored and build the models that would test all the possible combinations to determine the best of all.

References

[1] Voss EA, Boyce RD, Ryan PB, van der Lei J, Rijnbeek PR, Schuemie MJ. Accuracy of an automated knowledge base for identifying drug adverse reactions. J Biomed Inform. 2017 Feb;66:72-81. doi: 10.1016/j.jbi.2016.12.005. Epub 2016 Dec 16.

[2] Andrew M Wilson, Lehana Thabane and Anne Holbrook. Application of data mining techniques in pharmacovigilance. Br J Clin Pharmacol. 2004 Feb; 57(2): 127–134.

Coping with IT Downtime in Hospitals

Jackline BAMDEG^a and Thomas SCHMIDT^{a, 1}

^a*The Maersk Mc-Kinney Moller Institute, University of Southern Denmark*

Abstract. Downtime of information systems is a universal challenge faced by health care institutions. In this poster, we present the findings of a scoping review of how hospital organizations and their staff deal with downtime, and how coping can be grouped into three strategies; 1) Increasing communication, 2) Analog fallback, and 3) Restricted redundant systems. Our findings point to the importance of customizing coping mechanisms for individual healthcare institutions, and designing systems that empower users to deal with downtime.

Keywords. Health Information Systems, Downtime, Coping, Failure

1. Introduction

Anecdotal stories about system breakdowns and information technology (IT) downtime are frequent in health care. More rigorous investigations into the incidence of downtime has found that almost all healthcare institutions have experienced at least one unplanned downtime event, and that for the majority of these events the duration exceeded eight hours. Despite contingency planning and strategies for downtime handling, organizations still struggle to ensure that end-users of IT react sensibly during prolonged incidents. The aim of this work is to present an overview of coping mechanisms for handling unexpected downtime in hospitals. We focus on systems that can be fully, or partially, substituted with improvised solutions during the episodes.

2. Method

The study was conducted as a scoping review based on the search phrase ("technology failure" OR downtime) AND ("electronic health record" OR EHR OR "Information system" OR "electronic medical record" OR "Patient administration system" OR Paperless OR "Emergency care information systems" OR "Hospital Information Systems"). We searched Pubmed, Engineering Village, and Cinahl. Included papers was thematically analyzed to identify general coping categories.

3. Results

A total of 51 studies were assessed in full text, resulting in the inclusion of 13 studies for the final analysis. From the thematic analysis we structured the coping mechanisms

¹ Corresponding Author, Thomas Schmidt, Centre for Health Informatics and Technology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark.; E-mail: schmidt@mmmi.sdu.dk

into three categories: 1) *Increased communication* which sought to facilitate the lack of coordination offered by the IT system, 2) *Analog fallback* to capture the clinical information and knowledge needed for processing and treatment of patients in alternate ways, and 3) *Restricted redundant fallback* systems which still utilized IT but without access to the entire suite of functionality. The findings of our analysis are distilled into Table 1.

Table 1. Overview of results

Category	Coping mechanism	Affected system	Cause
Increased communication	Phone , Alternative working procedures, Email, Command center, Mass communication, Designated mediators	EHR, CPOE, LIS, HIS	System upgrade System launch, Hardware failure, Cyber-attack
Analog fallback	Printed labels, White board, Supplementary Paper Forms, Notebooks, Emergency Toolkit, Printed Ad-hoc logs	EHR, CPOE, CDSS, LIS, HIS	System upgrade, System launch, Hardware failure, Cyber-attack Natural disaster
Restricted redundant systems	Use of offline computers, Read-only systems	EHR, Medication administration	Hardware failure

The majority of downtime cases involve Electronic Health Records (EHR) or Hospital Information Systems (HIS), but also order entry systems (CPOE), clinical decision support (CDSS), and laboratory systems (LIS) were implicated. The coping mechanisms were coupled to the type of system affected.

4. Discussion

Given the steady improvement of hard- and software technology, one might expect that the risk of downtime would diminish over time, or even be an outdated issue. Yet, as illustrated by anecdotes as well as literature, this is not the case. During breakdowns, organizations either resolve to utilize broadcasting of information, or appointing designated carriers of information. Fallback systems are either improvised or designed in advance. In some cases, institutions were aware of the growing disparity between electronic or paper-based documentation and consequently prepared an entire emergency toolkit. From our initial navigation of the field, it is evident that grey literature of all kinds actually include a richer source of cases and experience. Given enough time and resources, there is most likely a lot of learning to be distilled from a systematic approach to collection and analysis of these stories.

5. Conclusion

The three coping strategies identified in this review, and the frequency by which they have been utilized, are indicative of the continued need for having fallback systems in place to handle lack of access to the IT systems. Although there are a number of research instruments available for gauging the impact of downtime, there is a need to harvest the experiences described in grey literature.

Creating and Implementing a Professional Development Program to Improve Organizational Capacity for Health Analytics in a Public Health Setting

Mariano AIZPURUA^{a,1}, Tomás GALLUZZI^a, Santiago ESTEBAN^a.

^a*Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina.*

Abstract. The implementation of an Electronic Medical Record (EMR) in the public healthcare system of Buenos Aires City began in June 2016. AGISE is a government department that assists in health decision-making processes by providing timely and quality data. In this study we designed a professional development program to improve organizational capacities for health analytics.

Keywords. professional development, training program, health analytics.

1. Introduction

The implementation of an Electronic Medical Record (EMR) in the public healthcare system of Buenos Aires City began in June 2016 as a central component of a Health Information System. The management of the large amounts of data that it generates is a complex, interesting, and vital challenge for the sustainability and development of this system and organization.

AGISE is a government department within the Health Ministry of the city of Buenos Aires, whose objective is to assist in health decision-making processes by providing timely and quality data to people of all decision levels. Through qualified professionals with heterogeneous profiles, AGISE aims to form a multidisciplinary work team within the public health sector. Current challenges in achieving this include finding new people who are able to adapt to desired profiles, identifying any gaps between the experience of the potential hires and the needs of the organization, and the difficulties in competing with the labor market due to less competitive salary usually offered by government institutions. The aim of this study is to describe the design and implementation of a comprehensive development program including a training curriculum for new members of the team and a continuous training program for the stable staff; to encourage a dynamic and collaborative analytical community in the public health sector in the context of new information and communication technologies (ICT).

¹ Corresponding author. E-mail: maizpurua@buenosaires.gob.ar.

2. Methods

The design of a competency-based educational program began with a review of the literature to rely on the best practices in professional development programs[1][2]. Expected analytic profiles were defined, with roles and tasks to be assigned. Training needs were assessed through semi-structured interviews with team members. As a result and according to the organization's mission, general and individual training and developmental goals were defined following the SMART criteria (Specific, Measurable, Achievable, Realistic, Timely). Transferable experience, skills and available resources were identified and analyzed to find out what the team was lacking. An action plan and timeline were developed, and its execution began in June 2019. Evaluation tools were designed to assess our progress and revise accordingly.

3. Results

The chosen training methodology is focused on competencies, with distributed learning following the principles of participation, repetition, relevance, transfer and feedback [3][4]. We selected priority contents for the curriculum of new members and organized them in an intensive two-week self-administrable package. For the continuous training of the stable team, we selected contents and organized them into seven thematic units distributed in an annual schedule. Small group projects, talks with experts, hands on workshops, theoretical classes, book clubs, and virtual courses were some of the modalities applied. For each thematic unit the programmed activities were scheduled in a shared calendar, their materials and resources stored in an accessible and indexed repository, and a forum was created to share questions, doubts, and experiences. We developed tools for both, evaluation and monitoring, to make it possible to adapt and optimize the curriculum's contents, as well as consider including new topics.

4. Conclusion

The implementation of a professional development plan that guides the initial training of new members and the continuous training of the team is a vital resource for the growth and development of organizations. Monitoring is necessary to assess progress and review accordingly.

References

- [1] Tremblay MC, Deckard GJ, Klein R. Health informatics and analytics - building a program to integrate business analytics across clinical and administrative disciplines. *J Am Med Inform Assoc.* 2016;23(4):824–828. doi:10.1093/jamia/ocw055
- [2] Miller SD, Stablein P, Syed J, et al. Evaluation of a Training Program to Improve Organizational Capacity for Health Systems Analytics. *Appl Clin Inform.* 2019;10(4):634–642.
- [3] Miguel Diaz, Mario de. n.d. *Modalidades de Enseñanza Centradas En El Desarrollo de Competencias.* Edited by Universidad de Oviedo. Ministerio de educación y Ciencia de España.
- [4] William B. Werther, Keith Davis. n.d. *Administración de Recursos Humanos. El Capital Humano de Las Empresas.* Mc Graw Hill.

Cross-Mapping Study of Nursing Practice Terms from a Brazilian Hospital Database

Luciana Schleder GONÇALVES^{a,1} Dayane ANDREATTA^b and Fernanda Karoline SCHAMNE^c

^a*Nursing Undergraduate and Graduate Program Professor, UFPR, Brazil*

^b*Nursing Undergraduate Program Student, UFPR, Brazil*

^c*Master in Nursing Science Program Student, UFPR, Brazil*

Abstract. Objective: To perform a cross-mapping study between nursing documentation from a Electronic Health Record (EHR) of a Brazilian hospital and four nursing terminology systems. Method: Descriptive, quantitative and retrospective study. Results: The EHR presents a total of 247 anamnesis / physical examination terms, 75 nursing diagnoses, and 205 nursing interventions terms to nurses make their choice. Conclusion: Authors strongly recommend that attempts in building more usable and friendly EHRs for clinical care practice documentation consider classification systems structures in their development, to ensure complete, safe, evidence-based and comparable registries.

Keywords. Nursing documentation. Nursing Informatics. Electronic Health Record. Terminologies. Cross-mapping.

1. Introduction

The Nursing Process (NP) is composed of interrelated steps that provide quality to nurse's clinical thinking and enable continuity of care in healthcare settings [1]. Anamnesis and physical examination data are gathered in its first stage and are essential to measure the resolution of care [2]. Nursing diagnosis (ND) is a clinical judgment about problems of an individual, family or community, over which nurses can intervene favoring their quality of life. This judgment is based on the nurse's assessment of the history, complaints, social and physiological data of the patient to whom nursing care is being delivered [3]. Nursing interventions (NI) are sets of actions that coordinate the "doing" of nursing care, and should be based on observation and evaluation of problems presented by the individual, family or community [1].

¹ Corresponding Author, Address: Rua Lothario Meissner, 632 Bloco Didático II, Departamento de Enfermagem, UFPR. Curitiba-PR-Brazil. E-mail: lualevale@gmail.com

In order to organize the specific language of nursing, classification systems have emerged, which allow the unification of concepts relevant to this area, considering their wide scope and performance in the fields of teaching, management, research and care [3].

2. Method

In order to cross-map anamnesis and physical examination, diagnoses and nursing interventions terms available in the hospital's EHR with four nursing standardized classification systems, a descriptive and retrospective study of quantitative approach was performed.

This study is part of a project entitled 'Artificial Intelligence and the management of sepsis patients: its influence on decision making, nursing interventions and outcomes', approved by the Health Sciences Sector of the Federal University of Paraná (UFPR, Brazil) Institutional Review Board (IRB) in May 2018, under registration number 84602218.8.0000.0102 and approved by the the co-participating IRB under registration number 84602218.8.3001.0098.

3. Results

From the EHR list of options, 247 terms referred to the description of anamnesis and physical examination, 75 to ND and 205 to NI. The match rates range from 58% to 100%, depending on the stage of the nursing process and the classification system mapped.

4. Conclusions

Care documentation is a challenge in current healthcare practice - it is a major cause of burnout among clinicians and nurses, and it is also mandatory for the continuity of care, visibility of nursing profession contribution and for legal and billing matters.

In this study a cross-mapping experiment is presented which reflects the importance of translational research in health and nursing informatics. Scientific knowledge, here presented by robust classification systems, provide the foundation for data, information, knowledge and wisdom structures that contribute to highlight clinical reasoning and evidence of healthcare practices. Authors strongly recommend that attempts in building more usable and friendly EHRs for clinical care practice documentation consider these structures in the development, to ensure complete, safe, evidence-based and comparable registries.

References

- [1] Horta WA. Nursing: theory, concepts, principles and process. Rev. Esc. Enf. USP. [Internet]. 1974 [cited 2019 Jul 2]; 8(1): 7-15. Available from: <http://www.scielo.br/pdf/reecusp/v8n1/0080-6234-reecusp-8-1-007.pdf>.
- [2] Santos N, Veiga P, Andrade, R. Importance of history and physical examination for the care of nurses. Rev. Bras. Enferm. [Internet]. 2011 [cited 2019 Jul 2]; 64(2): 355-8. Available from: <http://www.redalyc.org/pdf/2670/267019461021.pdf>.
- [3] Carvalho EC, Cruz DALM, Herdman TH. T. H. Contribution of standardized languages for knowledge production, clinical reasoning and clinical Nursing practice. Rev. Bras. Enferm. [Internet]. 2013 [cited 2019 Jul 2]; 66(esp): 134-41. Available from: <http://www.scielo.br/pdf/reben/v66nspe/v66nspea17.pdf>.

Data Integration Approaches for Representing Stem Cell Studies

Irena PARVANOV^a and Joseph FINKELSTEIN^{1a}

^a*Icahn School of Medicine at Mount Sinai, New York, NY, USA*

Abstract. The aim of this study was to examine existing methods for sharing results of stem cell research via online data repositories. To identify the relevant repositories, a PubMed search was conducted using standard MeSH terms which was followed by a web-based search of relevant databases. The search yielded 16 databases created between 2010 and 2019. The review of databases identified 35 major rubrics and their sub-rubrics organized in a five-module system. Data integration approaches were characterized by three domains (common data elements, data visualization and analysis tools, and ontology mapping) which varied widely across the databases. Current state of stem cell data integration lacks reproducibility and standardization. Standardization of data integration approaches for representing stem cell studies is necessary to facilitate data sharing.

Keywords. Data integration, data sharing, stem cells, standardization

1. Introduction

An exponential expansion of stem cell research results in voluminous data set aggregation [1]. However, effective approaches for intelligent integration and sharing of these resources have not been established. The aim of this study was to examine existing methods for sharing results of stem cell research via online data repositories.

2. Methods

Stem cell repositories were identified by using “Databases, Factual AND Stem Cells” MeSH terms in PubMed and web search for stem cells databases created between 2010 and 2019. All identified databases were manually reviewed and characterized.

3. Results

Overall, we located 14 academic databases indexed in PubMed and two databases not reflected in peer-reviewed literature (Table 1). The databases varied widely in data elements, analysis tools, search functionality and update methodology. Their features were summarized in 35 rubrics including names, lineage, surface markers, transcription factors as well as types of detection assays and clinical trial information.

¹ Corresponding author: Email: Joseph.Finkelstein@mountsinai.org.

Diverse rubrics were organized into a five module system based on Minimal Information About a Cellular Assay for Regenerative Medicine (MIACARM) [2] including Project, Assay, Source cell, Experimental technology, and Data.

Table 1: Stem cell databases

Database Tittle	Link
SyStemCell	http://lifecenter.sgst.cn/SyStemCell/
CODEX	http://codex.stemcells.cam.ac.uk/
ESTOOLS Data@Hand	https://ruoho.uta.fi/wp/estools/
Stem Cell Discovery Engine	http://discovery.hsci.harvard.edu/
StemFormatics	https://www.stemformatics.org/
LINCS	http://lincsportal.ccs.miami.edu/dcic-portal/
LifeMap Discovery	https://discovery.lifemapsc.com/
StemMapper	http://stemmapper.sysbiolab.eu/
ESCAPE	http://www.maayanlab.net/ESCAPE/index.php
SKiP Stemcell Knowledge and Information Portal	https://skip.stemcellinformatics.org/en/
hPSCreg	https://hpscreg.eu/
CellNet	http://cahanlab.org/macellnet.html
StemCellNet	http://stemcellnet.sysbiolab.eu/
HSC-explorer	http://mips.helmholtz-muenchen.de/HSC/
CORTECON	http://cortecon.neuralsci.org/
ESCD	https://biit.cs.ut.ee/escd/

4. Discussion

Data integration approaches were characterized by three domains (common data elements, data visualization and analysis tools, and ontology mapping) which varied widely across the databases. Current state of stem cell data integration lacks reproducibility and standardization. Standardization of data integration approaches for representing stem cell studies is necessary to facilitate data sharing. Consensus on common data elements and data exchange standards will facilitate data integration approaches in stem cell-based research and promote research transparency and reproducibility.

References

[1] Weissman I. Stem cell research: paths to cancer therapies and regenerative medicine. JAMA. 2005;294(11):1359-1366.

[2] Sakurai K, Kurtz A, Stacey G, Sheldon M, Fujibuchi W. First Proposal of Minimum Information About a Cellular Assay for Regenerative Medicine. Stem Cells Transl Med. 2016;5(10):1345-1361.

Data Provenance Standards and Recommendations for FAIR Data

Malte-Levin JAUER^{a,1} and Thomas M. DESERNO^a

^a*Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and
Hannover Medical School, Braunschweig, Germany*

Abstract. This article reviews the main characteristics of five widely used data provenance models and recommendations. We suggest a set of six provenance properties that should be satisfied by any provenance model as a basis for further implementation of provenance mechanisms, supporting the findable, accessible, interoperable and reusable (FAIR) principles for both, research and health data.

Keywords. Data provenance, FAIR data, Metadata, Research data, Health data

1. Introduction

In health research, data capture and data quality varies strongly. Therefore, information on data provenance is needed along the whole processing pipeline [1]. This includes the generation of persistent identifiers (PIDs) to make the data findable and accessible and is crucial to reuse data. Therefore, providing data provenance information is a mandatory step towards findable, accessible, interoperable and reusable (FAIR) data [2].

2. Methods

We consider five provenance standards identified within the FAIR4Health project [3]. A widely used provenance model is the W3C PROV-DM data model [4]: an acyclic directed graph, consisting of nodes “entity”, “activity”, and “agent”. Recommending specific provenance items, the DataCite International Consortium developed a metadata scheme in 2009 [5]. It stresses assignment of digital object identifiers (DOIs) and includes six domain-agnostic mandatory properties. In 2016, a domain-specific extension to the DataCite metadata schema for health was presented: the ECRIN Clinical Research Metadata Schema [6]. It includes information on the source study, associated consent and access details. The Research Data Alliance endorsed 14 recommendations of the Working Group Data Citation (WGDC) [7] targeting reproducibility of data used in experiments and studies. Therefore, persistent identifiers have to be generated in a query-based manner, so that data views can be cited and retrieved by re-executing the query. As a result of the Data Quality Collaborative (DQC), Kahn et al. [8] proposed 20 data quality and provenance recommendations. They especially highlight that each transformation of the source data has to be documented, including data cleansing values.

¹ Corresponding Author: Malte-Levin Jauer, Peter L. Reichertz Institute for Medical Informatics, Mühlenpfordtstr. 23, 38106 Braunschweig, Germany; E-mail: malte-levin.jauer@plri.de.

3. Results

We extracted the following list as minimal “fit for use” requirements for provenance model (Table 1). Check-marks indicate, which recommendation(s) support these items.

Table 1. Comparison of the different provenance recommendation sets.

Criteria	DataCite	ECRIN	WGDC	DQC
<i>Persistent identifier (PID)</i> : Each data object is assigned a unique, persistently stored identifier. Ideally, a DOI is assigned.	✓	✓	✓	✗
<i>Data origin</i> : The project or event that generated the data.	✓	✓	✗	✓
<i>Data creator</i> : A person or institution to be credited for.	✓	✓	✗	✓
<i>Data timestamp</i> : The time of dataset creation/modification.	✓	✓	✓	✗
<i>Data versioning</i> : Each transformation result of the data object is stored. Earlier versions are retrievable.	✓	✓	✓	✓
<i>Query PID</i> : If (sub-)sets of data are generated or cited, the query is stored with a persistent ID for reproducibility.	✗	✗	✓	✓

4. Discussion

The present work has identified six minimal criteria from the given provenance overview, implementable using the PROV-DM data model. The feasibility of these items will be investigated in the FAIR4Health project’s demonstrators.

Acknowledgements

This work was performed in the framework of FAIR4Health project. FAIR4Health has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement number 824666.

References

[1] Baum B, Bauer CR, Franke T, Kusch H, Parciak M, Rottmann T, et al. Opinion paper: data provenance challenges in biomedical research. *it - Information Technology*. 2017;59(4):191–196.

[2] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016 Mar;3:160018.

[3] FAIR4Health. D2.2. Functional design of the FAIR4Health platform and agents. Internal report. 2019.

[4] Moreau L, Missier P, editors. PROV-DM: the PROV data model [Online]. 2013 Apr 30 [cited 2019 Dec 21]. Available from: <https://www.w3.org/TR/2013/REC-prov-dm-20130430/>

[5] DataCite Metadata Working Group. DataCite metadata schema documentation for the publication and citation of research data v4.3 [Online]. Version 4.3. 2019 [cited 2019 Dec 21]. Available from: <https://schema.datacite.org/meta/kernel-4.3/>

[6] Canham S, Ohmann C. A metadata schema for data objects in clinical research. *Trials*. 2016 Nov;17(1):557.

[7] Rauber A, Asmi A, van Uytvanck D, Proell S. Data citation of evolving data: recommendations of the working group on data citation (WGDC) [pdf]. RDA WG Data Citation. 2015 Oct [cited 2019 Dec 21]. Available from: https://www.rd-alliance.org/system/files/RDA-DC-Recommendations_151020.pdf

[8] Kahn MG, Brown JS, Chun AT, Davidson BN, Meeker D, Ryan PB, et al. Transparent reporting of data quality in distributed data networks. *EGEMS (Wash DC)*. 2015 Mar;3(1):7.

Data Visualisation in Midwifery: The Challenge of Seeing what Datasets Hide

Bridget J DALEY^{a,1}, Evangelia KYRIMI^b, Kudakwashe DUBE^{c,d}, Norman E FENTON^b, Graham A HITMAN^a, and Scott MCLACHLAN^{b,d}

^a*Genomics and Child Health, Blizard Institute, Queen Mary University of London (UK)*

^b*Risk and Information Management, Queen Mary University of London (UK)*

^c*School of Fundamental Sciences, Massey University (NZ)*

^d*Health informatics and Knowledge Engineering Group (HiKER)*

Abstract. Information visualisation is transforming data into visual representations to convey information hidden within large datasets. Information visualisation in medicine is underdeveloped. In midwifery, the impact of different graphs on clinicians' and patients' understanding is not well understood. We investigate this gap and its potential consequences.

Keywords. data visualisation, midwifery

1. Introduction

Data visualisation is transforming data, information and knowledge into visual representations that convey meaningful patterns and trends hidden within large datasets [1]. Ineffective data visualisation negatively impacts clinical care and patient safety [2]. Effective visualisation mitigates comprehension, interpretability and data navigation issues [3]. Data visualisation in clinical medicine is underdeveloped [4]. This is especially true for midwifery. Finding appropriate ways to visualise data is challenging [5], and depends on objectives, context and audience [5]. Studies confirm the effectiveness of data visualisation. This paper investigates how data visualisation could be better applied in midwifery.

2. Method

A multi-database search was conducted using the terms “data visualisation” and “midwifery”. Duplicates, irrelevant papers and those which spoke of data visualisation but failed to present any form data visualisation in were removed. This resulted in a collection of just 10 papers for use in this review.

¹ Corresponding Author, E-mail: b.j.daley@qmul.ac.uk

3. Results

Our review identified that midwifery uses extremely limited and primitive data visualisation methods. Bar charts [6] and line graphs [7] dominated, with only one work [8] using any graphical representation. No works addressed visualisation methods for midwifery. Midwives are not being exposed to approaches that could improve their understanding and assimilation of clinical data, and accordingly lack awareness of approaches to and benefits of data visualisation.

4. Discussion

Direct patient care will continue to be the driving concern, but how midwives engage with patient data and research must evolve in the era of big data [9, 10]. Many consider data science as the next evolution for midwifery, yet few midwives understand what data visualisation is and the impact it may have [9, 10]. Increasingly, effective information visualisation may hold the key to understanding complex data in electronic health record datasets. Midwives must embrace data visualisation and the knowledge it can convey.

5. Conclusion

We have investigated the use of data visualisation within the domain of midwifery. Research is needed to assist midwives who work with or present data to do so using appropriate data visualisation. Without this, midwives will continue to be constrained by the challenge of seeing what datasets hide.

References

- [1] Ware, C. (2004). *Information Visualisation: Perception for Design*, 2nd Ed. San Francisco, USA: Morgan Kaufmann Publishers for Elsevier.
- [2] Pao, D., Stevens, J., Lockton, D., & Weinstein, N. (2018, July). Electronic Medical Records: Provotype visualisation maximises clinical usability. In EVA.
- [3] Carpendale, M. S. T. (2003). Considering visual variables as a basis for information visualisation.
- [4] IoM. (2011). *Health IT and Patient Safety: Building Safer Systems for Better Care*. Retrieved from: http://www.nap.edu/catalog.php?record_id=13269.
- [5] Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science* (New York, N.Y.), 333(6048), 1393–400. <http://doi.org/10.1126/science.1191181>.
- [6] Gao, F., Kihal, W., Le Meur, N., Souris, M., & Deguen, S. (2016). Assessment of the spatial accessibility to health professionals at French census block level. *International journal for equity in health*, 15(1), 125.
- [7] Vivian-Taylor, J., Sheng, J., Hadfield, R. M., Morris, J. M., Bowen, J. R., & Roberts, C. L. (2011). Trends in obstetric practices and meconium aspiration syndrome: a population-based study. *BJOG: An International Journal of Obstetrics & Gynaecology*, 118(13), 1601–1607.
- [8] Jönsson, A., & Bogers, L. Data-driven visibility: maternal bodies. *AIREA: Arts and Interdisciplinary Research*, 1, pp 63–79.
- [9] O'Connor, S. (2018). Big data and data science in Health Care: What Nurses and Midwives need to know. *Journal of Clinical Nursing*, 27, pp 2921–2922.
- [10] Peters, M. (2018). *Nursing Research: From integration with clinical practice to Future Science*. Australian Nursing and Midwifery Journal, 3326(2).

Deep Learning Approach for the Development of a Novel Predictive Model for Prostate Cancer

Mohaimenul ISLAM^{a,b}, Hsuan-Chia YANG^{a,b}, Phung-Anh NGUYEN^{a,b}, Yu-Hsiang WANG^{a,b}, Tahmina Nasrin POLY^{a,b}, Yu-Chuan (Jack) LI^{1a,b}

^a*Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan.*

^b*International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan.*

Abstract: We developed a deep learning approach for accurate prediction of PCA patients one year earlier with minimal features from electronic health records. The area under the receiver operating curve for prediction of PCA was 0.94. Moreover, the sensitivity and specificity of CNN were 0.87 and 0.88, respectively.

Keywords: Prostate cancer, deep learning, convolutional neural network, electronic health record.

1. Introduction

Prostate cancer (PCA) is the most commonly diagnosed cancer in men and the fifth leading cause of death globally [1]. Earlier stratification of patients at high risk of PCA offers an immense potential to decrease morbidity and mortality. Prostate-specific antigen (PSA), however, is a widely used and reliable test that is currently applied in screening the patients for correct identification of PCA patients. We, therefore, developed a novel prediction model using CNN algorithm with minimal features from HER that may help to stratify high-risk patients before a diagnosis of PCA. To our knowledge, it is the first study to develop a prediction model of identifying PCA patients one year earlier.

2. Methods

Two million population data between January 01, 1999 and December 31, 2013 was randomly selected to conduct this current study. All the patients who were ≥ 20 years old and diagnosed with prostate cancer retrieved from the database between January 01, 1999 and December 31, 2013. PCA patients were confirmed by using ICD-9 codes and record in the Registry of Catastrophic Illness Patient (RCIP), a subpart of the NHI database.

¹ Corresponding author Email: jaak88@gmail.com / jack@tmu.edu.tw

Patient to be included in the RCIP database must have pathological confirmation and treatment history of PCA. We then checked the total number of prescription and length of prescription for those patients. Afterwards, 7-days clinical information of all patients was summed up continuously. A total of 157 grid was constituted of 1095-days patients' information. We developed CNN model with a 5- fold cross-validation technique in this study. The area under the receiver operating curve, Sensitivity, specificity was used to measure the performance of our model.

3. Results

A total of 4,071 PCA and 16,284 controls were included in our current study. The mean age of case and control patients were 70.0, and 47.5 years, respectively. All the patients diagnosed with PCA were male. The average number of diagnosis in patients with PCA were higher than controls (44 vs 24) when we considered PCA and non-PCA patients one year before index date. However, over the three years, the average number of diagnosis in PCA and non-PCA patients was 41.23 and 22.66. The mean of AUC of the CNN model in the minimal electronic health record feature was 0.94, with sensitivity of 0.87 and specificity of 0.88 at the threshold value of 0.4.

4. Discussion & Conclusion:

In this current study, we focused on predicting PCA patients one year ahead from the minimal feature of electronic health records by the widely used CNN model. However, the CNN approach has shown high sensitivity and specificity for predicting PCA with the minimal feature of EHR. Based on these current findings, it is obvious that deep learning prediction model could provide physicians immense opportunity to improve patient outcomes through earlier PCA identification and subsequent follow-up high-risk PCA patients. Having provided timely treatment to patients may help to improve survival rate, minimize treatment costs and reduce risk of mortality [2]. We developed a novel deep learning risk prediction model for the prediction of PCA with minimal features of EHR. Our model, which includes patient demographics, comorbidities, co-medications, can be used to identify patients at high risk of PCA one year earlier.

References

- [1] Rawla P. Epidemiology of Prostate Cancer. *World journal of oncology* 2019;10(2):63.
- [2] Cordon-Cardo C, Kotsianti A, Verbel DA, Teverovskiy M, Capodiecici P, Hamann S, et al. Improved prediction of prostate cancer recurrence through systems pathology. *The Journal of clinical investigation* 2007;117(7):1876-83.

Defining a Standardized Information Model for Multi-Source Representation of Breast Cancer Data

Miguel PEDRERA^a, Pablo SERRANO^a, Ana TERRIZA^a, Jaime CRUZ^a,
Carolina VARELA^a, Ana SALAMANCA^a, Noelia GARCIA^b, Carmen SALGADO^a,
Consuelo SANZ^a, Lucía HERNÁNDEZ^a, Montse ALEMANY^a, Adolfo MUÑOZ^c
^a*Hospital Universitario 12 de Octubre, Madrid, Spain.*

^b*Instituto de Investigación Sanitaria Hospital 12 de Octubre, Madrid, Spain.*

^c*Instituto de Salud Carlos III, Spain.*

Abstract. This work aims to define a standardized information model for representation of multiple data sources in breast cancer. A set of data elements has been identified using ICHOM Breast Cancer as the minimum data set and adapting it to the needs of Hospital Universitario 12 de Octubre. With this, an information model has been defined according to ISO 13606 and SNOMED CT standards.

Keywords, EHR, Information model, Standardization, Breast cancer, ISO 13606, SNOMED-CT, ICHOM.

1. Introduction

Advanced uses of health information require standard data that allows its representation without loss of meaning [1], essential in a multidisciplinary condition such as breast cancer. This work aims to define an information model for the standardized representation of data in breast cancer domain.

2. Methods

First, the set of data elements for breast cancer domain was identified using ICHOM Breast Cancer [2] as the minimum data set. Second, this data set has been reviewed and adapted to the needs of Hospital Universitario 12 de Octubre [3]. Finally, based on this specification, an information model has been defined according to ISO 13606 [4] and SNOMED CT [5].

3. Results

Table 1 shows the number of data elements that have been identified, comparing them with the number of elements defined by ICHOM Breast Cancer.

Table 1. Data elements identified by categories and compared with ICHOM Breast Cancer

Data category	ICHOM BC	Data Element Set
Clinical	47	71
Pathological Anatomy	17	84
PROM	104	104
Total	168	259

Table 2 shows an example of standardized data element using the reference model of ISO 13606 and SNOMED CT.

Table 2. Data element standardized according to ISO 13606 and SNOMED CT standards

Component	Data-type	Cardinality	SCT Code	SCT Term
Element	Coded Value	1	371480007	Tumor site
Value	-	-	181134008	Entire upper outer quadrant of breast
Value	-	-	181136005	Entire lower outer quadrant of breast
Value	-	-	181132007	Entire upper inner quadrant of breast
Value	-	-	181135009	Entire lower inner quadrant of breast
Value	-	-	205042000	Entire central portion of breast
Value	-	-	265780004	Entire nipple

4. Conclusions

Data set proposed by ICHOM Breast Cancer has been used as the minimum data set. It has been necessary to increase the number of data elements from 168 to 259 to adapt it to the needs of the Hospital (see Table 1).

The combined use of ISO 13606 and SNOMED CT has allowed the standardized definition, structure and meaning, of previously identified data elements. This will allow advanced uses of these data with full meaning.

Acknowledgment

This work has been elaborated within the Research Projects PI15/00321, PI18/00981 and PI18CIII/00019, funded by Instituto de Salud Carlos III.

References

[1] Muñoz, A., Romero Gutiérrez, A., Marco Cuenca, G., Abad Acebedo, I., Caceres Tello, J., Sanchez Madariaga, R., ... & Maldonado Segura, J. A. (2013). Manual práctico de interoperabilidad semántica para entornos sanitarios basada en arquetipos.

[2] Ong, W. L., Schouwenburg, M. G., Van Bommel, A. C., Stowell, C., Allison, K. H., Benn, K. E., ... & Ganz, P. A. (2017). A standard set of value-based patient-centered outcomes for breast cancer: the International Consortium for Health Outcomes Measurement (ICHOM) initiative. *JAMA oncology*, 3(5), 677-68.

[3] Hospital Universitario 12 de Octubre (2019). Retrieved from <http://www.comunidad.madrid/hospital/12octubre/>.

[4] Serrano, P., Moner Cano, D., Sebastián, T., Maldonado Segura, J. A., Navalón, R., Robles Viejo, M., & Gómez, Á. (2009). Utilidad de los arquetipos ISO 13606 para representar modelos clínicos detallados. *RevistaeSalud. com*, 5(18), 100-110.

[5] Donnelly, K. (2006). SNOMED-CT: The advanced terminology and coding system for eHealth. *Studies in health technology and informatics*, 121, 279.

Design of a Supportive Transfer Robot System for Caregivers to Reduce Physical Strain During Nursing Activities

Christian KOWALSKI ^{a,1}, Pedro ARIZPE-GOMEZ ^a, Conrad FIFELSKI ^b,
Anna BRINKMANN ^b and Andreas HEIN ^b

^aOFFIS Institute for Information Technology, Oldenburg, Germany

^bCarl von Ossietzky University Oldenburg, Oldenburg, Germany

Abstract. The number of people in need of long-term care is rising and personnel scarcity is already foreseeable. The shortage of caregivers is further increased due to early retirement attributed to the major health burden when working at high speed with heavy lifting. Since nursing staff in many cases work beyond their physical strain limits during routine activities at the bed and existing systems do not counteract the current trend, we investigate with the present work whether the concept of a collaborative robotic support system can contribute to the physical relief of the nursing staff to make it possible to fall below the physical strain limits.

Keywords. robotics, human-robot collaboration, computer vision, camera calibration

1. Introduction

As the worldwide population is aging, the demand for caregiving personnel is increasing. In addition to this shortage of qualified personnel, the problem of physical overload, which results mainly from the high working speed and heavy lifting, prevails within the health care system. This overload in turn leads to an inability to work or an early retirement of the personnel [1], thus resulting in an even greater shortage of available staff. The biomechanical overload during routine nursing activities has been previously measured and described in detail [2]. In that work, the compressive force on the lumbosacral disc of the spine of caregivers during the execution of nine different types of nursing activities (e.g. moving the patient to the bed's head end) was estimated using a biomechanical model. In most cases, the estimated lumbar load values are not within acceptable limits. Consequently, technologies to reduce the physical load must be developed. While conventional support systems, such as patient lifts, have the drawbacks of slow patient transfer and limited usability, robotic solutions may have the potential to cooperatively support the nurse during physically demanding activities. In this work, we focus on the research of such a system in order to identify its limits.

¹Corresponding Author: Christian Kowalski, OFFIS Institute for Information Technology, Escherweg 2, 26121 Oldenburg, Germany; E-mail: christian.kowalski@offis.de.

2. Method

In order to provide comprehensive support for the movement of patients, an actuated bed is to be used. For the actual support at the bed, we chose the LBR iiwa 7 R800 model from Kuka, which makes safe human-robot collaboration possible due to its sensitive joint torque sensors [3]. To bring the components together, an intermediate construction was manufactured for the actuated bed on which the robot can be mounted. For interaction with the caregiver or the patient, additional sensors are required to detect the robot's environment. For this particular case, a visual system based on multiple 3D cameras is used to detect any unwanted collision at an early stage.

3. Results & Discussion

The result of the conceptional design can be seen in Figure 1. We decided to test the system's limits by using it in a scenario where the patient (rescue dummy, 80 kg) is turned to the side, then the caregiver places the hand on the patient's thigh for the transfer towards the head end of the bed. A marker on the nurse's hand is detected by the vision system and functions as the reference for the robot goal. At this predetermined position, the robot applies additional pressure with a predefined force so that the nurse supported and hence physically relieved during the process.



Figure 1. Final setup consisting of an actuated bed, mounted robotic manipulators and a Microsoft Kinect V2 vision system. The right side depicts a potential supporting scenario using the robot.

4. Conclusion

It was possible for us to realize a robotic assistance system mounted at an actuated bed with a visual detection system, which is able to support physically during predefined nursing activities. Next, we plan to record all regular physically demanding nursing activities executed by two cooperating nurses in order to make meaningful derivations for the operation of the robot on the basis of this visual data.

References

- [1] Glaser, J., & Höge, T. (2005). Probleme und Lösungen in der Pflege aus Sicht der Arbeits- und Gesundheitswissenschaften. Baua
- [2] Jäger, M., Jordan, C., Theilmeier, A., Wortmann, N., Kuhn, S., Nienhaus, A., & Luttmann, A. (2012). Lumbar-load analysis of manual patient-handling activities for biomechanical overload prevention among healthcare workers. *Annals of occupational hygiene*, 57(4), 528-544.
- [3] Vysocky, A., & Novak, P. (2016). Human-Robot collaboration in industry. *MM Science Journal*, 9(2), 903-906.

Detecting Severe Incidents from Electronic Medical Records Using Machine Learning Methods

Kazuya OKAMOTO ^{a,1}, Takashi YAMAMOTO ^b, Shusuke HIRAGI ^a,
Shosuke OHTERA ^a, Osamu SUGIYAMA ^c, Goshiro YAMAMOTO ^a,
Masahiro HIROSE ^d, and Tomohiro KURODA ^a

^a *Division of Medical Information Technology and Administration Planning,
Kyoto University Hospital, Japan*

^b *Patient Safety Unit, Kyoto University Hospital, Japan*

^c *Preemptive Medicine & Lifestyle-Related Disease Research Center,
Kyoto University Hospital, Japan*

^d *Faculty of Medicine, Shimane University, Japan*

Abstract. The goal of this research was to design a solution to detect non-reported incidents, especially severe incidents. To achieve this goal, we proposed a method to process electronic medical records and automatically extract clinical notes describing severe incidents. To evaluate the proposed method, we implemented a system and used the system. The system successfully detected a non-reported incident to the safety management department.

Keywords. Safety management, supervised machine learning, medical records

1. Introduction

In order to prevent medical accidents at hospitals, it is important to grasp those events at early stages that lead to severe medical accidents and then take appropriate actions based on the events. These events, which are defined as incidents, are usually reported by medical staff to the safety management department, that is responsible for the prevention of medical accidents in hospitals. The reports are analyzed by the department in order to decide which measures should be implemented by the clinical staffs. This mechanism is called an incident reporting system [1].

One of the main issues related to incidents is a long delay or even lack of reporting [2,3]. It is especially important to detect those cases since they affect the safety management departments ability to respond properly.

The goal of this research is to design a solution to detect these non-reported incidents, especially severe incidents. To achieve this goal, we develop methods to process electronic medical records and automatically extract clinical notes describing severe incidents. The extracted notes are treated as incident candidates which are shown to the safety management department for further analysis.

¹Corresponding Author: Kazuya Okamoto, Kyoto University Hospital, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto, 606-8507, JAPAN; E-mail: kazuya@kuhp.kyoto-u.ac.jp.

2. Methods

We develop methods to process electronic medical records and automatically extract clinical notes describing incidents of injection by using the SVM based technique [4].

First, we manually label a training set of clinical notes into two categories based on whether they include a severe incident report or not. Next, by morphological analysis, the training set is separated into words and arranged in a vector space using single words as the axes. Then, the SVM creates the machine learning models from the arranged training set with labels.

Finally, based on the created machine learning models, the SVM extracts positive clinical notes which are estimated as clinical notes with incident reports. The extracted notes are treated as incident candidates which are shown to the safety management department for further analysis.

3. Experiments

Using the proposed method based on the linear kernel SVM, we implemented an incident candidate reporting system. To evaluate the system, we asked a staffs of the safety management department to judge whether extracted incident candidates were incidents or not.

4. Results

The system extracted 121 incident candidates from the 294,731 clinical notes. 34 of them were judged to include severe incident reports by the staff of the medical safety department, which means that the precision is 28.1%. 31 of them were related to sudden critical changes of patients' conditions; other two were related to incidents during surgeries; the other was related to a fall from a bed. Moreover, several sudden critical changes among the detected incidents were not reported actually.

5. Conclusions

In this research, we aimed to establish a method to extract incident candidates from clinical notes in order to detect non-reported severe incidents. In addition, we implemented a reporting system that presents incident candidates extracted by using the proposed method. The system successfully detected a non-reported incident to the safety management department, thus our goal was achieved.

References

- [1] T.K. Nuckols, D.S. Bell, S.M. Paddock, L.H. Hilborne LH: Contributing factors identified by hospital incident report narratives, *Qual Saf Health Care* **17**(5) (2008), 368–72.
- [2] M. Hirose, S.E. Regenbogen, A.A. Gawande: Lag time in an incident reporting system at university hospital in japan, *Qual Saf Health Care* **16** (2007), 101–104.
- [3] S.M. Evans, J.G. Berry, B.J. Smith, A. Esterman, P. Selim, J.O. Shaughnessy, M. DeWit, Attitudes and barriers to incident reporting: a collaborative hospital study, *Quality & Safety* **15** (2006), 39–43.
- [4] B.E. Loser, I.M. Guyon, V.N. Vapnik: A Training Algorithm for Optional Margin Classifiers, *Proc. of the 5th Annual ACM Workshop on Computational Learning Theory* (1992), 144–52.

Developing a Mobile App for Patients to Monitor Medical Record Changes Using Blockchain

SungJun PARK^a, MinDong SUNG^a, Sungjae JUNG^b, Eunsol LEE^b,
Jaehoon LEE^{c,d}, Yu Rang PARK^a

^a*Department of Biomedical Systems Informatics Yonsei University
College of Medicine*

^b*Medibloc Inc.*

^c*Intermountain Helathcare, SaltLake City, UT*

^d*Department of Biomedical Informatics, University of Utah, Salt Lake City, UT*

Abstract. According to recent revisions to medical laws in Korea, changes to electronic medical records are to be documented. To do so, however, a transparent system with which to store original documents and changes thereto is needed. The transparency and immutability of blockchain records are the key characters of blockchain technology. Employing these characteristics, we developed an application with which to monitor changes of medical records using blockchain.

1. Introduction

Although medical records are personal to the individual patient, patients themselves often face difficulties with gaining access to their medical records, compared to medical staff. If a patient would like to determine whether their medical records have been changed, they must request all medical records at a hospital. Also, according to Korean medical law, accessing electronic medical records is to be documented.[1] An objective and transparent system with which to preserve the history of revisions to original medical data is needed. Accordingly, we developed a system that allows patients to monitor changes to their medical records using blockchain technology, which is known for its transparency and immutability.[2]

2. Method

To begin, we categorized information contained within medical documents into three risk levels: medically important and incapable of being modified; medically relevant, but capable of being modified; and medically unrelated. Unlike existing blockchain transactions, transactions used in private blockchains do not include the recipient's address, allowing for a separate variable in the blockchain to distinguish patients. The transaction structure of a blockchain block comprises four items: the Owner, Topic, Key, and Value. The Key and Value items hold four additional items: Hash, DocumentUri, PreviousHash, and Modified. The Hash item represents the unique ID of a medical record, and DocumentUri indicates the address where the medical record is stored. PreviousHash reflects the previous version of the current record of the Hash. When a medical record is entered for the first time, the PreviousHash value is entered as Null. Finally, the Modified item indicates how many changes have been made to the

medical record. This application was developed with Android Studio using Java version 1.8 and utilizes the published blockchain API.

3. Results

This application monitors the creation of medical records, changes thereto, the risk of changes, and the creation and modification timestamps through the blockchain. The processes of the application through the blockchain are shown in Figure 1.

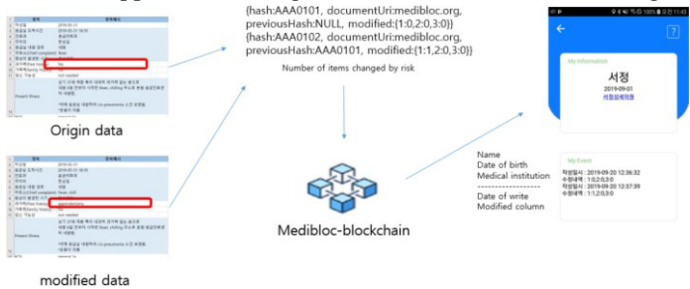


Figure1. How the application works

On the left side, the origin data and modified data are recorded in the blockchain in the same text format as the center. So you can see how many your medical records been changed or not in the application through the medibloc panacea-core API.

4. Conclusion

In this study, we established a monitoring system that allows patients to be aware of changes to their medical records using blockchain technology. Using the application, patients can be informed about the status of their medical record as soon as the record is created or changed. Adopting this system will allow for more transparent management of medical records and potentially more trust among patients in their doctors. There have been cases of building electronic medical records⁴ or personal health records⁵ with blockchain. However, only the construction of a blockchain system that stores information has been discussed, and most proposed systems have been established for only the storage and sharing of medical records. This study is the first to develop an application for monitoring changes to electronic medical records, and a prospective study on its usability is being planned. The blockchain itself is decentralized, but it's not a perfect DApp. Due to it gets its value from the database to access the app. In addition, there are still Smart contracts security issues. Therefore, DApp is discussion and limitation.

References

[1] Para. 4 of Art. 23 of Medical Law
[2] Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing Blockchains for Efficient Health Care: Systematic Review. J Med Internet Res. 2019;21(2):e12439–12. doi:10.2196/12439.

Developing and Piloting TB Medication and Refilling Reminder System in Ethiopia

Kassahun DESSIE GASHU ^{a,1}, Fedlu NURHUSSIEN ^b, Adane MAMUYE ^b, Kassahun ALEMU GELAYE ^c and Binyam TILAHUN ^a

^a *Department of Health Informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia*

^b *College of Informatics, University of Gondar, Gondar, Ethiopia*

^c *Department of Epidemiology and Biostatistics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia*

Abstract. This project followed waterfall model to develop a context tailored and user-friendly reminder system to support home based TB treatment. piloting was conducted on expert and patient levels. We developed a flexible and user-friendly web-based system that allows medication and refilling reminder. Participants agreed on its usefulness despite infrastructure and personal related challenges.

Keywords. Phone reminder, medication, refilling, Ethiopia

1. Introduction

Non-adherence to TB treatment worsen during continuation phase [1]. Mobile phone messaging showed promising output in supporting home-based medication [2]. Most of mHealth Apps run with smart technology but smartphone ownership is low in Africa and usability remained challenging due to poor tailoring and engagement of end-users during development process [3]. This project aimed to develop and test a web-based reminder system for home-based TB medication using both basic and smart phones.

2. Methods

The project is part of an ongoing randomised controlled trial study in Northwest Ethiopia, registered in PACTR201901552202539. The project applied a waterfall model development process and piloted at expert and patient levels. We used NetBeans, Glass Fish, Java Persistence API, java platform Enterprise Edition, JavaScript and MySQL for frontend and backend development, and Serreta modem to handle SIM cards. Piloting, involved 2 IT experts, 3 health information technicians and 4 TB focal persons to evaluate interface and layouts. To assess usefulness and challenges, 23 adult patients on continuation phase enrolled and received medication and refilling reminder for 2 months.

¹ Corresponding Author,; E-mail: kassahundessie@gmail.com

3. Result

We developed a flexible and user-friendly web-based reminder system. The main features include: registration, reminder scheduling for medication and refilling (Fig.1). About, 20(87%) of patients received all reminders. Of them, all replied the system was useful. A respondent replied: *“Rural people like me are busy and forget pills. This system helped me lot.”* (42Y/O, M). Participants reported: weak mobile network, power inconsistency, medication time variation (4(17%) patients follow suggested time).

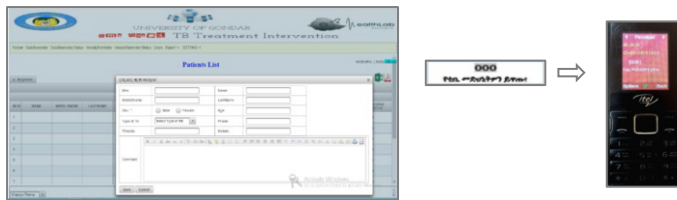


Figure 1. Patient registration and cell phones based text (Amharic) and graphic reminder for TB treatment

4. Discussion

The current project developed a context tailored, user-friendly and flexible web-based reminder system for home-based TB treatments. Globally, various generic mHealth systems developed, however, usability and sustainability often determined based on contexts [4]. In resource limited setting where no reliable ICT infrastructure, low digital literacy and low access to smartphone and needs to consider those factors during development [5]. The mobile cellular subscriptions shown fastest growth 74% in Sub-Saharan Africa including 51% in Ethiopia [6]. Therefore, development of mHealth technologies need to assess local requirements and tailored to the context. As a limitation, this reminder system works only with text and simple graphic messages.

5. Conclusion

A context tailored reminder system could be usable for home-based TB treatment in low resource settings. The randomised controlled study could address its efficacy.

References

- [1] Adane, A.A., et al., Non-Adherence to Anti-Tuberculosis Treatment and Determinant Factors among Patients with Tuberculosis in Northwest Ethiopia. PLOS ONE, 2013. **8**(11): p. e78791.
- [2] WHO, mHealth: new horizons for health through mobile technologies: 2nd global survey on eHealth. Global observatory for eHealth Series, 2011, Geneva: WHO. 102.
- [3] Boulos, M.N.K., et al., Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online journal of public health informatics, 2014. **5**(3): p. 229-229.
- [4] Achaliya, P., An Android Based Medication Reminder System Based on OCR Using ANN. 2013.
- [5] Lester, R., et al., Mobile phone SMS for adherence support and care of patients with TB infection: Evidence and opportunity. Journal of Clinical TB and Other Mycobacterial Diseases, 2019. **16**.
- [6] World Bank, Mobile cellular subscriptions, ITU, World Telecommunication/ICT Development Report and database <https://data.worldbank.org/indicator/IT.CEL.SETS.P2>. 2016

Development of a Graph-Based Database for Ovarian Cancer Symptoms

Young Ji LEE^{a,b1}, Harleigh NIYU^c

^a*School of Nursing, University of Pittsburgh, USA*

^b*Department of Biomedical Informatics, University of Pittsburgh, USA*

^c*Department of Mathematics and Computer Science, Duquesne University, USA*

Abstract. Uncontrolled cancer-related symptoms severely affect the quality of life of cancer patients. This study leveraged symptom data written in the online health communities (OHCs) and developed a symptom knowledge database for cancer patients. Further study is needed to apply this pilot study to the larger OHCs data.

Keywords. graph-based database, text mining, consumer health informatics

1. Introduction

Uncontrolled cancer- and treatment-related symptoms not only severely affect the quality of life, but also decrease the overall survival rate of cancer patients [1]. In the Internet era, many patients share their symptom experiences on online health communities (OHC) that is not attainable from clinical data [2]. However, it is difficult to utilize the information written in the OHCs without a standardized approach to discover and manage information. Thus, there is a need to develop a pipeline to utilize this wealth of information for cancer research and clinical practice in the future.

The overarching goal of the study is to develop a symptom knowledge database for cancer patients. As a pilot study, we will focus on ovarian cancer (OvCa). OvCa is a strong example for understanding the complex challenges of managing symptoms because OvCa patients have a wide range of tumor types and many face a lifetime of ongoing active disease management due to high recurrence rates [3].

2. Methods

Data was collected from the OHC sponsored by the American Cancer Society: Cancer Survivors Network (<http://csn.cancer.org/forum/132>). We extracted posts that were posted between January, 2006 and March, 2016 (n=50,626). As a pilot study, we randomly selected 100 postings that described “nausea” and manually annotated them in four concepts: symptoms, treatments, results and other medications. After annotation, each concept was entered as a node in the Neo4j (<https://neo4j.com/>), an open-source graph-based database, allowing us to visualize data in an easily understood manner.

¹ Corresponding Author, School of Nursing, University of Pittsburgh, 3500 Victoria Street, Pittsburgh, USA 15213; E-mail: leeyoung@pitt.edu.

3. Results

Our current database includes data from 88 users either an OvCa patient or caregiver of OvCa patient and 266 nodes. We have identified 27 symptoms including “nausea”, 10 cancer-related treatment (e.g. Taxol, cisplatin) and 4 different effect of treatments. We also identified 6 co-existing medical conditions and 31 non-cancer related medications such as Tylenol. Current database includes 5 types of relationship between nodes: “diagnosed_with”, “has”, “resulted_in”, “takes”, and “underwent”. Figure 1 showed an example of graph-based database.

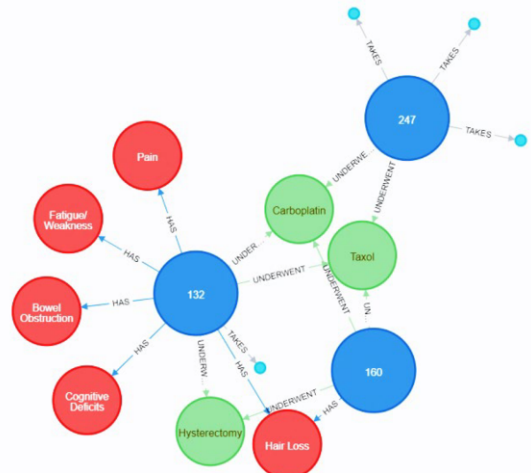


Figure 1. Example of nodes and relationships described in Neo4j

4. Discussion and Conclusion

We found that OHCs can capture more symptom related information compared to the current symptom assessment checklists. Furthermore, it is crucial to understand relationship between symptoms in a symptom cluster since cancer symptoms exist as a cluster of symptoms rather than a single symptom. Further study will apply natural language processing to automate the annotation process for larger datasets and capture additional factors (e.g., age, cancer stage, race, and author type) from the OHCs.

Our pilot study showed the potential of OHCs in the nursing research and clinical practice to understand cancer symptoms from the patients’ perspectives that have not been captured in the clinical setting. We believe that this will guide clinicians and researchers about to consider when designing interventions for managing symptoms.

References

- [1] E. Basch, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA* 318 (2017): 197-8.
- [2] Y. Lu, et al. Health-related hot topic detection in online communities using text clustering. *PLoS One* 8 (2013): e56221
- [3] H.S. Donovan, et al. Representations of fatigue in women receiving chemotherapy for gynecologic cancers. *Oncol Nurs Forum* 32 (2005):113-6.

Development of a Personalized mHealth Program for Cancer Symptom and Complication Management

Lixin SONG,^{a1} Eno IDIAGBONYA,^a Kevin SU,^a and Juliet IDIAGBONYA^a

^aSchool of Nursing & LCCC University of North Carolina-Chapel Hill

Abstract. Based on care need assessment of colorectal and bladder cancer (CBC) patients with newly created ostomies (stomas) and caregivers, and with the input of the Wound, Ostomy, and Continence Nurses (WOCNs), we have developed an mHealth complication and symptom monitoring and self-management program, titled *PRISMS*.

Keywords. cancer, patient-reported outcome, symptom management

1. Introduction

Ostomy creation is one of the major treatment options for colorectal and bladder cancer (CBC), the third and fifth most common cancers in the U.S., respectively.¹ Up to 80-90% of CBC patients with new ostomies experience dehydration, infection, and other symptoms and complications after being discharged from hospitals.^{2,3} The symptoms and complications are associated with reduced quality of life (QOL),⁴ depression and anxiety, and increased ER visits, hospital readmission,^{3,5} and high costs of care.³ This study aimed to develop an mHealth intervention to enhance personalized supportive care and improve QOL for cancer patients and caregivers during care transition from ostomy creation professional care in the hospital to self-management at home.

2. Methods

We conducted mixed-method formative research to identify the care needs among post-discharge CBC patients and their caregivers. Based on analysis of the data abstracted from Reddit (www.reddit.com/) (n=206 patients and caregivers) and from a large survey among caregivers through the Bladder Cancer Advocacy Network (BCAN) (N=559), and qualitative interview of WOCNs, we have developed a web-based mHealth complication and symptom monitoring and self-management program, titled PRISMS.

3. Results

3.1. Social media analysis. We found most of the posts were written by children of patients (n=62, 55%), followed by patients themselves (n=33, 29%). Patients looked for support from others who have had similar experiences (peer support) (n=33, 29%) and talked about their physical symptoms and psychological/emotional struggles (both n=29, 26%). Family members expressed psychological/emotional struggles (n=42, 37%), asked

¹ **Corresponding author:** Lixin Song, PhD, RN, FAAN, University of North Carolina-Chapel Hill, Carrington Hall, Chapel Hill, NC 27599-7460. Tel: 919-966-3612. Fax: 919-843-9900. Email: lsong@unc.edu

for advice on how to provide emotional and behavioral support to the patient (n=24, 21%), and sought peer support (n=15, 13%). Other post-treatment supportive care information needs included future quality of life (n=19, 17%), treatment side effects (n=11, 10%), and medical/financial resources (n=10, 9%).

3.2. BCAN survey. The majority of caregivers in this cross-sectional online survey were spouses or partners, female, and college-educated or higher. Mean age of caregiver was 61 years (21-85). Although caregivers needed support for physical well-being, more than two thirds needed emotional support. Web-based programs and peer support were most preferred intervention format

3.3 PRISMS development. Based on these results, a literature review, and the input of WOCNs (N=3), we identified the following needs of CBC patients with ostomies and caregiver: continuous access to post-discharge supportive care at home, knowledge of symptom and complication monitoring and management, and support from peers and professionals. We developed PRISMS to address these needs, to reduce symptoms and complications, and ultimately, to improve the QOL for both patients and caregivers and reduce unnecessary ER use and readmission for patients. PRISMS' features include:

- a) Supportive care needs for the most common post-surgical symptoms and complications, e.g., fatigue, dehydration, skin care, sleep issues, and emotional distress;
- b) Scalable self-management program used in home settings that can provide self-monitoring and self-care guidance, communication with peer patients and caregivers, and access to health care professionals if needed.
- c) Knowledge and skills training for symptom and complication management, moderated online forum for peer and professional support, and other resources.
- d) Integration of PRO and objective data from wearable devices to provide continuous monitoring of patients' symptoms and complications.
- e) Use of plain language to accommodate health literacy needs for this population that is mostly comprised of seniors: <https://www.plainlanguage.gov/>

3.4 *We have completed user testing of the PRISMS prototype among 10 nurses and physicians, as well as 10 CBC cancer patients with ostomies and their caregivers who have been extremely enthusiastic about the features of PRISMS and also provided useful information for refining PRISMS.*

3.5 *We will conduct a 3-arm pilot randomized trial for feasibility testing among 45 CBC cancer patients and their caregivers in the next 12 months.*

References

- [1]. American Cancer Society. *Cancer Facts & Figures 2018*. Atlanta 2018.
- [2]. Malik T, Lee MJ, Hari Krishnan AB. The incidence of stoma related morbidity – a systematic review of randomised controlled trials. *Annals of the Royal College of Surgeons of England*. 2018;100(7):501-508.
- [3]. Stitzenberg KB, Chang Y, Smith AB, Nielsen ME. Exploring the Burden of Inpatient Readmissions After Major Cancer Surgery. *Journal of Clinical Oncology*. 2015;33(5):455-464.
- [4]. Chongpison Y, Hornbrook MC, Harris RB, et al. Self-reported depression and perceived financial burden among long-term rectal cancer survivors. *Psycho-oncology*. 2016;25(11):1350-1356.
- [5]. Bell JF, Whitney RL, Reed SC, et al. Systematic Review of Hospital Readmissions Among Patients With Cancer in the United States. *Oncology nursing forum*. 2017;44(2):176-191.

Development of a Questionnaire to Measure Healthcare Providers' Satisfaction with Telemedicine Services

Esmée TENSEN^{ab,1}, Youri BROEKHUIZEN^{ab*}, Johan van BUGGENUM^{a*}, Leonard WITKAMP^b, Monique JASPERS^a, and Linda PEUTE^a

^a*Amsterdam UMC, University of Amsterdam, Department of Medical Informatics, Amsterdam Public Health Research Institute, Amsterdam, Netherlands*

^b*Ksyos Health Management Research, Amstelveen, Netherlands*

**Second and third author contributed equally*

1. Introduction

With the advent of technology in healthcare, healthcare providers (HCPs) are increasingly using telemedicine services that enable them to diagnose or treat patients by means of digital remote consultation of a specialist. Due to their shown cost-effectiveness and proven quality of care, telemedicine services have now permeated a wide variety of fields such as cardiology, ophthalmology, and dermatology. HCPs' satisfaction (i.e. needs, wishes and expectations) with these services in health practice is a primary factor in the usage, acceptance and sustainability while dissatisfaction with these services will reduce their willingness to use the system. However, a validated questionnaire focusing on HCPs' satisfaction with the use of integrated telemedicine services is lacking. We aim to develop such a tool. This validated instrument will capture individual perspectives on telemedicine services independently and confidentially. This abstract describes the steps followed to develop a first version of the questionnaire.

2. Method

Dutch experts from a telemedicine healthcare organization (Ksyos) participated in this project. Ksyos works with 12,000 HCPs; medical specialists, paramedics and GPs. First, an explorative literature search was conducted to acquire existing validated questionnaires on job-, communication-, and telehealth service satisfaction in healthcare. Two validated questionnaires, the staff satisfaction questionnaire used in the Basque health service and the telehealth usability questionnaire (TUQ), were explored to develop a list of potentially relevant constructs as input for our questionnaire [1,2]. Second, an appreciative inquiry of all revealed constructs was performed with a Ksyos domain expert and a human factor expert. Third, the questions were forward translated to Dutch by two researchers. Then, a focus group with account managers and a customer service employee was held to discuss the resulting constructs that would form the basis for the first version of the questionnaire. We used their input because they

¹ Corresponding author: Esmée Tensen, E-mail: e.tensen@amsterdamumc.nl.

have frequent contact with HCPs who use telemedicine services. During the focus group meeting consensus was reached by rating all constructs on relevance. Question formulation and applicability to the construct were likewise discussed. Questions could be added if construct seemed missing or changed if in need of rephrasing. Fourth, a pre-testing of the questionnaire took place among HCPs in two rounds. Eighteen HCPs received a PDF-document with all the questions (e.g., "I believe [*name telemedicine system*] suffices in fulfilling my wishes and expectations for telemedicine") and were also asked to review each question phrasing in the document and, for certain questions, the applicability to their context. In the end, the questionnaire was built in an online questionnaire tool (LimeSurvey) and an invitation link was sent to another 21 HCPs to check for technical problems with filling in the digital questionnaire and to measure the needed completion time.

3. Results

The two questionnaires from the literature resulted in a list of 20 potentially relevant constructs. After the appreciative inquiry, three construct themes were not applicable within the current research setting and deleted. The focus group led to textual and contextual changes to the phrasing of the questions. Nineteen of 58 questions were rated as 'not relevant', four questions were added and two questions were merged. Respectively 10/18 and 8/21 unique HCPs participated in two pre-testing rounds. Again, textual changes were made based on terminology, three questions were removed and four questions were added in this phase. Besides the background characteristics, the remaining 49 questions could be classified in seven themes: working conditions, training, pay, communication, organization and policy and strategy, interaction with the telemedicine system, usage of the telemedicine system. Average questionnaire completion time was 9 minutes (2nd round) and no technical problems occurred.

4. Discussion and conclusion

Positive contributions of telemedicine on quality and costs of care are essential and shown. Equally important is to understand if, why and what makes HCPs embrace these innovative care services. HCPs' satisfaction with telemedicine from different perspectives is ideally measured by use of a questionnaire. Such a tool may also allow telemedicine service benchmarking on quality criteria from the HCP perspective in different care fields. Based on a structured pre-testing approach, we developed a first version of a telemedicine service questionnaire. One of our limitations is that our current version of the questionnaire was mainly based on the store-and-forward context of the Ksyos system and therefore did not focus on direct video consulting services.

References

- [1] C. Matthies-Baraibar et al., Is organizational progress in the EFQM model related to employee satisfaction? *BMC Health Serv Res* **14** (2014), 1-10.
- [2] B. Parmanto, A.N. Lewis Jr., K.M. Graham, M.H. Bertolet, Development of the Telehealth Usability Questionnaire (TUQ), *Int J Telerehabilitation* **8** (2016), 3-10.

Development of a Smart e-Coach Recommendation System for Obesity

Ayan CHATTERJEE ^{a,1}, Martin W. GERDES^a and Santiago MARTINEZ^b

^aUniversity of Agder, Faculty of Engineering and Science, Grimstad, Norway

^bUniversity of Agder, Faculty of Health and Sports Sciences, Grimstad, Norway

1. Introduction

By 2030, 8 out of the 10 leading causes of death will be linked to risk conditions of ‘Lifestyle diseases’, irrespective of gender [WHO]. Obesity is one of the major lifestyle diseases that lead to other health conditions, such as cardiovascular disease, chronic obstructive pulmonary disease, cancer, diabetes type II, hypertension, and depression. e-Health monitoring has become increasingly popular providing ICT-based remote and timely care support to patients and healthcare providers. In our research, we are working on a smart e-Coach solution to perform regular monitoring of health and wellness parameters related to obesity. We are going to develop a smart e-Coach system utilizing ICT, IoT and AI technologies to provide individual behavioral recommendations aiming at a healthier lifestyle to prevent obesity.

2. Method

We have divided our project work into the following task groups based on the well-established design-science-research methodology: (1.) Systematic literature review; (2.) Design and development (feasibility study, data collection, development of a recommendation engine, e-Coaching through interaction); (3.) Trial run; and (4.) Model evaluation with performance parameters (goal achievement). The feasibility study includes (a.) What to measure? [independent variables (e.g., demographics) and dependent variables (e.g., weight)]; (b.) How to measure? [spending time to explore and find the right instruments and methods]; (c.) Type of data to be collected [personal, physiological, contextual and behavioral] and determination of data collection process; (d.) Policy for the recruitment of the participants; (e.) Preparation of consent form; (f.) Ensuring data security and privacy; (g.) Use of a standard framework for eHealth intervention and determination of its efficacy. This research is deeply focusing on **“What to coach” and “How to coach”, including** the design, development, testing and evaluation of observational evidence-based, context specific and individual obesity recommendations. We have targeted group size of participants to 50-99 in Norway within an age group of 18-40 (both male and female). Healthy participants will be recruited on voluntary basis following a physical condition check and a consent form

¹ Corresponding author E-mail: ayan.chatterjee@uia.no.

[approved by the Regional Komiteer for Medisinsk og Helsefaglig Forskningsetikk (REK, <https://helseforskning.etikkom.no/>)] handover process during recruitment.

3. Results

The project is in an early design phase with the focus on the data collection part. We have designed a solution architecture to collect data from participants and store them securely. We are working on the 'eHealth ontology design' that is another important part in our study to support the structuring and to establish interoperability between different heterogeneous system components and networks. In this poster presentation, we will demonstrate charts and diagrams related to 'Health and Wellness Data Collection Process', 'Type of Data to be Collected', 'Solution Approach', and 'eCoach Recommendation plan for Achieving Goal'.

4. Discussion

The analysis of the problem space has shown three research directions that we will follow in this research: (1.) Data acquisition and its security; (2.) Health risk prediction; (3.) Personalized e-Coaching. We have targeted three main problems to solve in this research project: (a.) Creation of a compact, intelligible abstraction from massive, unintuitive raw, unstructured observations for health and wellness data (e.g. sensors, user preferences, questionnaires.) using an eHealth ontology; (b.) Protection against illegitimate access to the system and EHR; (c.) Development of a digital e-Coach system to model meaningful, observational and empirical evidence-based, context specific, automatic, personalized recommendations to achieve health wellness goals, and to advise about physical activity, nutrition, medication, and other relevant factors for healthier lifestyle. The digital e-Coach system will capture physiological (blood pressure, heart rate, blood cholesterol, blood glucose, height, weight), contextual (time, location, environment, social), and behavioral (activity, nutrition, sleep) data over time from secure wearable sensors, customized questionnaires, interviewing, apps, feedback forms and minimal-invasive way. Time series analysis of collected data will help to determine day-by-day random variation in lifestyle, weekly variation, seasonal variation, variation in diet, meal pattern, true changes in diet over time, and behavior change. In the recommendation process, participants will be given personalized, evidence-based, contextual recommendations and motivation (related to activity, nutrition) for long-term goal evaluation to keep BMI in the range of 18.5 - <25.

5. Conclusion

We are working on the assumed hypothesis that an effective e-Coach mechanism can reduce the obesity risk with automatic generation of personalized recommendations. Integrating offline (human) coaching psychology to an e-Coach, selection of an effective recommendation plan and improving it further to reach the convergence criteria (overweight to normal weight) is a computationally hard problem that we have planned to solve in our research with eCoach behavioral (diet, activity) interventions.

Development of a Systematic Text Annotation Standard to Extract Social Support Information from Electronic Medical Records

Camila VOLIJ^{a,1} and Santiago ESTEBAN^a

^a*Family and Community Medicine Division, Hospital Italiano, Buenos Aires, Argentina*

Abstract. The influence of social relationships on mortality is comparable with that of standard clinical factors. In order to extract this information via supervised NLP algorithms, clinical notes have to be annotated manually. At present we did not find any systematic annotation standards to extract social support. In this work, we developed a systematic text annotation standard to detect social support in couples from electronic medical records.

Keywords. Social Support, Electronic Medical Record

1. Introduction

The influence of social relationships on mortality is comparable with that of standard clinical factors such as smoking, obesity and hypertension [1]. In the last few years, the National Academy of Medicine recommended that social support be documented in electronic medical records (EMRs) [2]. Social determinants of health, and particularly, social support is usually recorded in free text form within EMRs. In order to apply supervised natural language processing pipelines (NLP) to extract this information, texts have to be manually annotated. Nevertheless, we were unable to find any systematic annotation standards for social support comparable to other already existing standards (i.e. TimeML). Thus, in this work we aim to develop and test an annotation standard that allows to create supervised datasets in order to extract information on social support from life partners.

2. Methodology

First, based on Valtorta et al. [3] systematic review, we thoroughly reviewed 54 questionnaires on social support. Subsequently, a research committee of EMR specialists and users from our institution, evaluated the various questions and selected those that were considered, a priori, that could be answered via the EMR. Finally, we developed an

¹ Corresponding author;; E-mail: camila.volij@hospitalitaliano.org.ar.

annotation standard that allowed us to describe the social support provided by a life partner defining entities and exploring the relationships between them.

3. Results

After the literature search and the focus groups with professionals, we defined social support from life partners in two dimensions: structural and functional support. We operationalized structural support by means of five variables: 1) Is there a mention of a type of personal relationship status? (true/false); 2) Does the relationship status refer to the patient (true/false); 3) What type of relationship status is mentioned (married, widower, 'in a relationship', single, etc); 4) Is the patient currently in a stable relationship? (true/false); 5) Is the patient currently living with whomever she is in a relationship with? (true/false).

We operationalized functional support using four variables: 1) Is there a mention of active social support in the text (true/false); 2) Is the patient the recipient or the provider of social support?, 3) Is the mention of social support positive or negative?, 4) Type of social support (psychosocial, healthcare related, general care related).

We tested our annotation standard on a random sample of 2000 clinical notes that were previously filtered using a keyword-based filter (the filter had a positive predictive value of 81% and negative predictive value of 99% for detecting texts with mentions regarding social support from a life partner in an unfiltered random sample of clinical notes). 96% of those 2000 clinical notes had mentions on the patient's current relationship status. Of these, 75% described stable relationships and 8% of them, that the patient and their partner lived together. Regarding functional social support, 28% of the notes mentioned some type of social support. Of those notes 90% mentioned the patient being the recipient. Lastly, in 89% of these notes the support was positive.

4. Discussion and Conclusion

The systematics of text annotation presented in this work, although presented as an innovative and efficient way of detecting intra-social support, has as a limitation the construct itself being difficult to define and therefore to measure even with questionnaires well known. Defined variables and their categories may be used for the creation of systematic annotation of texts on family social support, intra-friends, within institutions, among others.

References

- [1] Holt-Lunstad J, Smith TB, Layton JB (2010) Social Relationships and Mortality Risk: A Meta-analytic Review. *PLoS Med* 7(7): e1000316. <https://doi.org/10.1371/journal.pmed.1000316>
- [2] Institute of Medicine. Capturing social and behavioral domains and measures in electronic health records: phase 2. Washing, DC: The national academies press; 2014.
- [3] Valtorta NK, Kanaan M, Gilbody S, Hanratty B. Loneliness, social isolation and social relationships: what are we measuring? A novel framework for classifying and comparing tools. *BMJ Open* 2016;6:e010799.

Digital Health Interventions for Diabetes Self- Management Education/Support in Type 1 & 2 Diabetes Mellitus

Dumisani NKHOM^a Charles Jenya SOKO^a Pierre BOWRIN^{tb} and Usman IQBAL^{a,b,c},

^a *Masters' Program in Global Health & Development, College of Public Health, Taipei Medical University*

^b *PhD Programs in Global Health & Health Security Dept., College of Public Health, Taipei Medical University*

^c *International Center for Health Information Technology (ICHIT), Taipei Medical University*

Abstract. Introduction. Diabetes Mellitus is a major global concern. DSME (Diabetes Self-Management Education and Support) helps improve outcomes. Over the years, newer technologies have delivered DSME warranting an update of evidence. **Objectives:** Describe current digital health interventions employed in DSMES and their effectiveness in improving glycemic control and quality of life of Diabetes Mellitus (DM) patients compared to usual care. **Methods:** Systematic review design and meta-analysis. We will search data from 2010 to 2019. The review will only include randomized control trials (RCTs) with digital health interventions (mobile health, social media, e-health) as the main intervention to deliver DSMES. **Expected Results:** We expect digital health intervention-led DSMES to improve target outcomes compared to usual care. This study will aggregate information on usage and challenges for these tools. It will also suggest the direction of effect of such tools in different populations.

Keywords. Diabetes Self-Management Education and Support,

1. Introduction

Diabetes Mellitus is a major global concern and a costly disease. Currently, Diabetes Mellitus is a top ten global killer of humans. One method to improve outcomes and reduce costs is Diabetes Self-Management Education and Support (DSMES). DSMES also helps in preventing occurrence of complications. Ideally, quality DSMES should be easily accessible, however, its availability varies with different health systems. Digital health interventions help to solve these challenges with varying success. In addition, continuous digital innovations warrant up-to-date evidence for digital health interventions in DSMES. Therefore, we aim to know: What digital health interventions are being used in DSMES to improve glycemic control, Diabetes knowledge and health-related quality of life (HrQoL) for people with T1DM and T2DM in the past 10 years? & What is the effectiveness of digital health interventions for DSMES in improving glycemic control, Diabetes knowledge and HrQoL as compared to usual care for people with T1DM and T2DM in the past 10 years?

2. Methods

Design will be systematic review and meta-analysis. The study framework and reporting will be done in accordance to Preferred Reporting Instrument for Systematic Review and Meta-Analysis (PRISMA). We will search literature from the following databases: PubMed, EMBASE, Cochrane Library, Web of Science and Scopus. The review will only include randomized control trials (RCTs) published between 2010 and 2019 comparing digital health interventions delivering DSMES to usual care focusing on glycemic control, HrQoL and Diabetes Knowledge. Studies targeting other populations apart from confirmed Type 1 and 2 DM patients will not be included. In this review, healthcare that does not include use of digital health interventions will be considered as usual care. Protocol is registered on PROSPERO CRD42019139884. Articles are going to collected into Endnote X9. Studies will be screened against the eligibility criteria when looking at the titles and abstracts. Full text review will be done by two reviewers with disputes resolved through consensus.

Table 1: Proposed study variables that may help explain effect of digital health interventions on outcomes

Population	Interventions	Study	Outcomes
Age, sex, sex ratio, ethnicity, sample size, country, education level, disease type, duration of disease	Digital category Description Length Frequency Follow-up time	Setting Design Theoretical framework Comparison Author Journal Year of publication	HbA1c HrQoL Knowledge Instruments used Scales used

We will adapt Cochrane Collaboration’s data extraction form. Missing information will be sought through emails. Differences will be resolved through consensus or consulting a third reviewer

1.1 Analysis

Variables will be summarized using descriptive statistics. We will use Cochrane Risk of Bias (ROB)-2 tool for parallel and cluster randomized studies. Assessment will be done independently by two reviewers, reaching consensus on differences. Random effects meta-analysis will be performed according to Diabetes type, digital health intervention route at various time points using Comprehensive Meta-Analysis Software Version 3 (Biostat, USA). Outcomes will be collected in one measuring unit or be aggregated using standardized mean difference. Qualitative analysis will assess acceptability & feasibility of digital interventions. Findings from subgroup analysis and meta-regression will help explain interaction of study variables with outcome estimates. The above will considered of great clinical importance. Publication bias will assess across studies.

2. Expected Results

We expect digital health intervention-led DSMES to improve target outcomes as compared to usual care. This study is expected to aggregate information on common usage and challenges for these tools. It will also inform us the direction of effect of such tools in different populations and environments. This study will be an addition to literature in support of Sustainable Development Goal 3 in the fight against non-communicable diseases and in ensuring universal health coverage.

Electronic Medical Records Virtual Course. Strategies for an Effective Training

Daniel A. RIZZATO LEDE^a Cintia D SPERANZA^a Florencia ALEGRE^a

Adriana CANTORA^a Emilio PANIZZA^a Ariadna POU^a

^a *Ministry of Health, Argentina*

Abstract. The Introductory Course to the Electronic Medical Records (EMR) is part of the Digital Health Strategy of the Ministry of Health's Universal Health Coverage (UHC). One of its principles is achieving interoperability between different electronic medical records systems. The healthcare workforce training in these subjects is essential. During 2018, six editions of the course were developed in the Moodle Virtual Health Platform with 850 approved participants. We analysed the previous experience to reformulate the learning materials and create new instances for pedagogical support.

Keywords. eLearning, Electronic Health Records, Change Management.

1. Introduction

According to the Ministry of Health (MoH) Resolution 189/2018, Digital Health is one of the three strategic principles to achieve Universal Health Coverage in Argentina. Interoperable EMR can store primary information during contact with the patient to build a national medical record with clinical, statistical and management utility.

Evidence shows that the implementation of any EMR is a complex process that may have to overcome multiple barriers: financial, technical, time, psychological, social and resistance to change [1]. Negative beliefs of social and psychological nature play a significant role in resistance to adopt EMR [2]. The potential of eLearning for health professionals is widely researched and reported [3]. The EMR online educational course for health professionals develops arguments to tear down usual adoption barriers, and challenges every student to become a champion of change to transform his own work practices. We analysed our previous online courses experience to reformulate the learning materials and enhance the pedagogical support.

2. Materials and Methods

Argentina has a decentralized federal organization divided into 24 jurisdictions. In its territory there are about 30,000 health centers of all levels of complexity that provide care for 45 million people. Around 60% of the population has fixed Internet access and there are 30 million users with mobile connectivity. Health workers are heterogeneous in digital literacy, but most of them interact with technological tools.

Taking into account the amount and distribution of potential students (3,000 for the first stage), the virtual self-administered course was planned to be available

periodically in the Moodle Virtual Health Platform of the MoH, at no charge. Flexibility is one of the features that makes virtual education particularly suitable for most citizens, mainly adults [4]. After 6 cohorts, 850 students passed their examinations. Quantitative and qualitative data was gathered from course enrollment, follow up, passing rates and feedback surveys. Following a technical review, the course was redesigned to ease usability and comprehension. New cohorts were also analyzed.

3. Results

The original course had 3 sections: 1) General aspects of EMR; 2) Privacy, Confidentiality and Security; 3) Interoperability. The new edition kept the same sections, with great didactic evolution in: Language (technicalities were avoided, explicit references to more complex concepts were included); Content (simplification of complex concepts focusing on the main idea); Applicability (examples were provided to enable correlation with professional practice); Audiovisual tools (short videos and animations with clear locution voice, infographics to synthesize theoretical notions); Interaction (an online forum to answer questions and promote student exchange); Exams (focus on global understanding of central ideas); Follow up (enrollment in an alumni forum to exchange local experiences).

Table 1. Number of enrolled and approved students during the last 3 cohorts of the online course (2019).

Number of students	7 th edition	8 th edition	9 th edition
Enrolled / Finished / Approved	397 / 182 / 163	561 / 242 / 220	388 / 179 / 157

Online surveys showed a trend towards better student perceptions. Table 1 shows the outcomes of the new course. Even though there was more than 50% dropout, there was a higher passing rate.

4. Conclusion

This course redesign aims to exploit the potential of e-Learning by identifying the better approach to teach about EMR in our context. The flexibility in space and time of online tools combined with proper contents favor the achievement of the strategy objectives. Enabling the MoH Moodle Virtual Health Platform for mobile devices could increase the benefits of this modality, including higher enrollment and retention.

References

[1] A. Boonstra, and M. Broekhuis, Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions., *BMC Health Serv. Res.* **10** (2010) 231. doi:10.1186/1472-6963-10-231.

[2] M.V. Giussi, A. Baum, F. Plazzotta, P. Muguerza, and F. Gonzalez Bernaldo de Quirós, Change Management Strategies: Transforming a Difficult Implementation into a Successful One., *Stud. Health Technol. Inform.* **245**(2017) 813–817.

[3] P. Delf, Designing effective eLearning for healthcare professionals, *Radiography.* **19** (2013) 315–320. doi:https://doi.org/10.1016/j.radi.2013.06.002.

[4] L. García Aretio, M. Ruiz Corbella, D. Domínguez Figaredo, *From long distance education to virtual educacion*, Ariel, Barcelona (2007), 80-86.

Electronic Medication Administration System Supports Safe Medication Administration

Jenni SANTAVIRTA^{a,1}, Anne KUUSISTO^b, Kaija SARANTO^c, Tarja SUOMINEN^d
and Paula ASIKAINEN^c
^{a, b, c} *Satakunta Hospital District*
^c *University of Eastern Finland*
^d *Tampere University*

Abstract. The purpose of this study is to describe nurses' views of what supports safe medication administration in the current electronic medication administration system. Data was collected at the turn of 2014–2015 and open-ended answers were inductively analyzed using content analysis. The system's usefulness, good usability, and the feature that there is extra information available on medications and the patient-specific information needed in medication administration are elements that support safe medication administration. The study identifies wide support for the electronic medication administration system in safe medication administration.

Keywords. Safety, medication administration, electronic medication administration system, nurse

1. Introduction

Prevention of medication errors and avoidable adverse events^{1,2} is an important part of patient safety¹ and nurses' work². The use of electronic medication administration system has been found to reduce the rate of medication errors in medication administration recording and their potential future risk³.

The purpose of this study is to describe nurses' views of what supports safe medication administration in the current electronic medication administration system.

2. Methods

This study is part of wider electronic health record implementation research. Data was collected from special health care in one hospital district in Finland with one two-part open-ended section of the back-translated Medication Administration System - Nurses Assessment Scale (MAS-NAS) (N=855). In the open-ended section, nurses were asked to write down their comments about the current medication administration system and

¹ Jenni Santavirta, Corresponding author, Sairaalan tie 3, 28500 Pori, Finland; E-mail: jenni.santavirta@satasairaala.fi.

the degree to which the components of the current system enables them to administer medication safely and professionally. In this study, we report answers that described support for electronic medication administration system (eMAS) in safe medication administration. The data was collected during one month at the turn of 2014–2015. At that time, the respondents used the eMAS partly, completely, or not at all. The eMAS included an electronic home medication chart, electronic ward medication chart, and electronic prescription. Open-ended answers were analyzed inductively using content analysis.

3. Results

A total of 123 nurses answered the open-ended section; 80 of them had used eMAS partially or completely. Twenty nurses described what supports safe medication administration in the current eMAS; 15 of them reported that they had used eMAS partially or completely. Twenty-eight simplifications were formed. The eMAS elements that support safe medication are the system's usefulness to user, system's good usability, and the feature that there is extra information available in the system of medications and the patient-specific information needed in medication administration. Usefulness of the system was described as clarity of information, automated warning alerts, and continuity of medication. Good usability of the system was described as good technical quality and ease of use. Extra information available in the system was described as seeing the information needed in medication administration, such as information on medication preparation. Availability of patient-specific information needed for medication administration was described as the opportunity to read the medication list and prescriptions being in the patient record.

4. Discussion and conclusion

Nurses identified elements in the eMAS that support safe medication. Elements that support safe medication were the system's usefulness, usability, and availability of information needed for medication. Because one of the most important parts of nurses' work is to administer medication safely², nurses' opinions of how the system supports safe medication administration should be taken into account. The study identifies that eMAS widely supports safe medication administration. Further development of the systems should ensure that the system is useful for nurses, its usability is good, and the information needed for medication administration is available for nurses.

References

- [1] L. Kim, C.H. Lyder, D. McNeese-Smith, L.S. Leach & J. Needleman, Defining attributes of patient safety through a concept analysis, *Journal of Advanced Nursing* **71**(11) (2015), 2490-2503.
- [2] E. Rohde & E. Domm, Nurses' clinical reasoning practices that support safe medication administration: An integrative review of the literature, *Journal of Clinical Nursing* **27** (2018), e402-e411.
- [3] N. Vicente Oliveros, P. Pérez Menendes-Conde, S. Martín-Aragón Álvarez & E. Delgado Silveira. Effect of an electronic medication administration record application on patient safety. *Journal of Evaluation in Clinical Practice* **23** (2017), 888-894.

Emotion Analysis Using Electrodermal Signals and Spiking Deep Belief Network

Nagarajan GANAPATHY¹ and Ramakrishnan SWAMINATHAN

*Non-Invasive Imaging and Diagnostics laboratory, Department of Applied Mechanics
Indian Institute of Technology Madras, Chennai, India - 600036*

Abstract. In this work, an attempt has been made to discriminate arousal and valence dimensions in Electrodermal Activity (EDA) signals using Spiking Deep Belief Network (SDBN). EDA signals having different arousal and valence dimensions are obtained from public online database. These signals are divided into equal parts and normalized using channel normalization. Later, signals are subjected to SDBN for event-related features and classification. A leave-one-out cross validation is used to investigate the classification performance. The result shows that the SDBN classifiers are able to discriminate the emotional states. The network yields better classification performance for emotional dimensions of arousal and valence. It appears that the proposed approach can be used to differentiate autonomic and pathological conditions.

Keywords. Emotion, Electrodermal Activity, Spiking Deep Belief Network,

1. Introduction

Emotions are the fundamental component of human social interactions. It is necessary for behavior, perceptions and decision making. Analysis and recording of Electrodermal Activity (EDA) signals are widely used for the classification of emotions. Deep learning architectures are used for the automated recognition of emotions using physiological signals [1].

2. Methods

For this study, EDA signals are considered from publicly available DEAP database [2]. These signals are digitized at sampling rate of 128 Hz, each one minute in length and are obtained from 32 subjects. These signals are preprocessed and divided into 60 equal segments. The segments are applied to Spiking Deep Belief Network (SDBN) for training. A two-layer 50-50 neurons SDBN architecture is used for this study. Weights at individual nodes are updated using Leaky Integrate-and-Fire (LIF) neurons. The activation probabilities of LIF neurons is identified using an accurate rate-based approximation. The standard dynamics of the LIF neurons is given by [3]:

$$\tau_m \frac{dv}{dt} = E_{res} - V_{mem} + M_r \left(\sum_{i=1}^n w_i \sum_{j=1}^{m_i} \delta(f_i - f_{t_j}) \right) \quad (1)$$

¹ Email:info.nagarajan@gmail.com

where τ_m is the membrane time constant, E_{res} represents the resting potential, and M_r depicts the membrane resistance, n represent the number of incoming synapses, m_i depicts the number of spikes arriving at that synapse, $\delta(ft)$ is a Dirac-delta function set as zero except for the firing times ft_{ij} of the i^{th} input neuron.

3. Results and Discussion

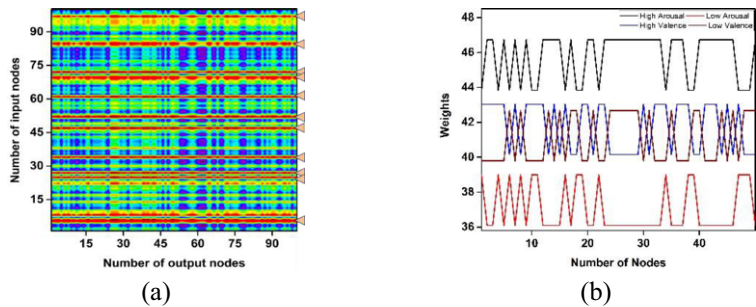


Figure 1. (a) The fine-tuned DBN weights in the second layer, and (b) A typical EDA signal class signatures generated by SDBN for classification

The visualization of fine-tuned DBN weights at the second layer is shown in Fig. 1(a). The critical features obtained in the second layer are highlighted using solid triangles. It is observed that the critical features are sparse and represents the abstract characteristics of various emotional states. The signatures of arousal and valence class generated by the SDBN for the classification are shown in Fig. 1(b). It is seen that the class signatures of arousal states are distinct from each other. In comparison to arousal class, valence class are also found to have different signatures. The classification performance of the proposed network are shown in Table 1. It is seen that the proposed approach obtained an average accuracy of 72.08% and 68.90% for arousal and valence states, respectively. It is also observed that the proposed approach obtained a precision of 74.70% for arousal class. In case of valence class, the proposed approach obtained the higher recall rate of 61.06%.

Table 1: Classification performance of SDBN architecture

Class	Accuracy	Precision	Recall
Arousal	72.08	74.70	83.23
Valence	68.90	50.96	61.26

References

[1] Greco, G. Valenza, L. Citi and E. P. Scilingo, “Arousal and Valence Recognition of Affective Sounds Based on Electrodermal Activity”, IEEE Sensors Journal, 2016, 17(3), 716-725.

[2] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “Deap: A database for emotion analysis using physiological signals”, IEEE Transactions on Affective Computing, 3(1), 18–31, 2012

[3] Vreeken, J. “Spiking neural networks, an introduction”. Institute for Information and Computing Sciences, Utrecht University Technical Report,2002, 8.

Evaluating the Use of a CPOE for Chemotherapy Protocols

Jimenez, C^a, de Matos Lima, S^a, Luna, D^a

^a *Department of Biomedical Informatics, Hospital Italiano de Buenos Aires, Argentina.*

Abstract. Chemotherapy drugs are one of the most common causes of serious and fatal medication errors, especially during prescribing, where computerized physician order entry (CPOE) take on importance. This study proposes the description of the post-implementation status of a CPOE in a highly specialized hospital between January and June 2018, among patients older than 18 years. Results: a total of 8835 protocols were indicated using the specific CPOE (93% use rate over all protocols) 91% completed the administration, 1.2% were rejected by pharmacy, and 6.8% was canceled. The most frequent cause of rejection by pharmacy and cancellation by oncologist was an inadequate dose. Most of the protocols indicated using the CPOE implemented, with a reject by pharmacy rate of 1.2%, indicates the utility of CPOE as an error prevention strategy.

Keywords. Computerized provider order entry, oncology, chemotherapy

1. Introduction

Chemotherapy drugs are characterized by low therapeutic index and significant toxicities at clinically prescribed doses, which, added to the vulnerability of patients with cancer, raises serious problems of drug safety [1]. Medication errors can occur at any stage of the chemotherapy process, especially at the prescription stage [2]. The most common source of error was within the order phase and that, compared with non-chemotherapy medication errors, could rise up to 48% more likely to be serious in nature [3]. The safety of chemotherapy has been a long-standing issue, particularly because is a complex process, with multiple steps and health professionals involved in the prescription, dispensing and administration of these agents [4][5]. CPOE is associated with a significant reduction in preventable adverse drug events [6] and focusing on chemotherapy, most studies find beneficial when compared to paper prescription [7].

2. Methods

A descriptive cross-sectional study was carried out in order to study the amount of oncological prescriptions made from the specific CPOE.

3. Results

During the study period, a total of 10475 protocols were indicated, of which 1380 were excluded from the San Justo center (other medical center in the network), 255 since they did not have an assigned sector and 5 protocols that were indicated in a pediatric area, giving a total of 8835 protocols. Each of the protocols corresponds to a patient. The CPOE was used in 93% of the oncological prescriptions. The remaining 7%

corresponded to indications on paper, drugs not approved by the Committee of the Cycle of Medication and drugs of research protocols. There was a total of 195 protocols loaded in the management system, of which 146 were indicated. There was an average of 1472 protocols per month. A total of 7602 (91%) protocols were finalized. Oncology pharmacy put in review 107 protocols (1.2%). The number of canceled protocols 602 (6.8%) includes both those canceled by medical initiative and those placed on review by pharmacy. In 41% (249) of the cases the oncologist entered a justification in free text at the time of cancellation. The most frequent reason was the need to modify the dose (33% - 82/249). There was 5% (13/249) of justifications with illegible content. Within the justifications charged by oncological pharmacy to put a protocol in the status "In review" (1.2%), 47% was related to an inadequate dose. All the justifications were complete regarding the reason and the instructions for correction by the physician.

4. Discussion

Numerous studies worldwide report the advantages of having a computerized physician order entry for prescribing chemotherapy protocols, supported by the evidence of providing patient safety, comparing with paper prescription [1]. The percentage of use of the CPOE was 93%, which indicates that most of the prescribed protocols used this system. Similarly, a study conducted in Switzerland where only 7% of the total prescriptions were manual and also highlight that the average error rate decreased 22 times [4].

5. Conclusion

Most of the protocols indicated used the CPOE for prescribing chemotherapy protocols implemented, with a rate of use greater than 90%. The rate of protocols intervened by the pharmacist of 1.2% indicates the utility of a CPOE as an error prevention strategy.

References

- [1] Elsaid KA, Garguilo S, Collins CM. Chemotherapy e-prescribing: opportunities and challenges. *Integr Pharm Res Pract.* 2015;4: 39–48.
- [2] Rahimi R, Kazemi A, Moghaddasi H, Arjmandi Rafsanjani K, Bahoush G. Specifications of Computerized Provider Order Entry and Clinical Decision Support Systems for Cancer Patients Undergoing Chemotherapy: A Systematic Review. *Chemotherapy.* 2018;63: 162–171.
- [3] Kukreti V, Cosby R, Cheung A, Lankshear S, ST Computerized Prescriber Order Entry Guideline Development Group. Computerized prescriber order entry in the outpatient oncology setting: from evidence to meaningful use. *Curr Oncol.* 2014;21: e604–12.
- [4] Voeffray M, Pannatier A, Stupp R, Fucina N, Leyvraz S, Wasserfallen J-B. Effect of computerisation on the quality and safety of chemotherapy prescription. *Qual Saf Health Care.* 2006;15: 418–421.
- [5] Hammond P, Harris AL, Das SK, Wyatt JC. Safety and decision support in oncology. *Methods Inf Med.* 1994;33: 371–381.
- [6] Nuckols TK, Smith-Spangler C, Morton SC, Asch SM, Patel VM, Anderson LJ, et al. The effectiveness of computerized order entry at reducing preventable adverse drug events and medication errors in hospital settings: a systematic review and meta-analysis. *Syst Rev.* 2014;3: 56.
- [7] Collins CM, Elsaid KA. Using an enhanced oral chemotherapy computerized provider order entry system to reduce prescribing errors and improve safety. *Int J Qual Health Care.* 2011;23: 36–43.

Evaluation of Mobile Phone Mortality Risk Score Applications Using Data from the Electronic Medical Records

Nino FIJACKO^{a,1}, Lucija GOSAK^a, Primož KOCKEK^a, Leona CILAR^a, Andrej MARKOTA^b, Gregor ŠTIGLIC^a

^aUniversity of Maribor, Faculty of Health Sciences, Maribor, Slovenia

^bUniversity Medical Centre Maribor, Maribor, Slovenia

Abstract. Various mobile phone apps in the form of medical calculators are available for different prognostic assessments, especially for patients in intensive care units. We performed a systematic review of mobile phone apps in online mobile phone stores to identify apps for mortality risk prediction in intensive care units. Out of 2737 potential mobile phone apps, we included 20 of them in the final content analysis. The most frequently used mortality risk model was Sequential Organ Failure Assessment also known as SOFA. The mobile phone apps were compared based on realistic electronic medical record data. The discrepancies were shown in patients with lower mortality rate. Our results show that this kind of mobile phone apps can be helpful to healthcare professionals and are appropriate for use in clinical practices in most cases.

Keywords. Critical care, mortality, prognostic models, m-health

1. Introduction

In intensive care unit (ICU), healthcare professionals frequently use different prognostic models (PMs) for the quantification of severe physiological derangements in patients. Many PMs are automatically calculated as part of the electronic medical record (EMR) and become a part of health care process. Other can be also used in various mobile phones as mobile phone applications (apps) [1]. In our study, we wanted to identify potentially suitable apps for illness severity prediction in clinical settings and to quantify the characteristics of these apps.

2. Methods

We performed a systematic review of apps in two mobile phone stores (Google Play Store and App Store (iOS)) based on PRISMA recommendations [2]. Using the predictive parameters from the most frequently used PM, we created two different fictional patients with different values using data obtained from the MIMIC III database [3].

¹ University of Maribor, Faculty of health sciences, Žitna street 15, 2000 Maribor, Slovenia; E-mail: nino.fjacko@um.si.

3. Results

Out of 2737, we included 20 apps into detailed analysis among which we identified 38 different PMs. The most commonly used PMs were SOFA and APACHE. The PMs required different types of values for the predictions, such as vital functions (i.e. arterial pressure), laboratory test results (i.e. creatinine) and demographics (i.e. age of the hospitalized patient). Values of both fictional patients were put into SOFA score apps where mortality scores in all apps were consistent, but deviations occurred in the interpretation of mortality, were not.

4. Discussion

Predicting the severity of illness of critically ill patients in the ICU is a comprehensive process that helps physicians anticipate the course of treatment. Based on the data obtained, we found that most apps provide the user with information on which PM was used, since all relevant apps contained information on the scientific research underlying their model. This indicates the quality of PMs, as the calculations in all PMs are uniform and accurate. Discrepancies have occurred in the interpretation of mortality itself, as individual apps interpret the obtained estimate differently. The difference is most obvious in patients with lower mortality rate.

5. Conclusion

In the future, apps for risk prediction will become an important part of healthcare, as they represent a faster and easier access to relevant health information and can be easily used with the patient, allowing us to make a more relevant assessment as the patient's condition can be monitored all the time.

Acknowledgments

This work was supported by the Slovenian Research Agency grant N2-0101.

References

- [1] Iorio-Morin, C., Fortin, D. & Blanchard, J., TBI prognosis calculator: a mobile application to estimate mortality and morbidity following traumatic brain injury, *Clinical neurology and neurosurgery* 142 (2016), 48-53.
- [2] Moher, D., et al., Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement, *PLOS Medicine* 6 (2009), e1000097.
- [3] Johnson, A., et al., MIMIC-III, a freely accessible critical care database, *Scientific data* 3 (2016).
- [4] Fijacko, N., Brzan, P. P. & Stiglic, G., Mobile Applications for Type 2 Diabetes Risk Estimation: a systematic review, *Journal of Medical Systems* 39 (2015), 124.

EVENT : A Concept Developed to Improve the Use of EHR by the Nurses

Alain JUNGER

^a *Centre hospitalier Universitaire Vaudois (CHUV – UNIL,
Lausanne, Switzerland)*

Abstract. Context and purpose. The University Hospital of Lausanne has been deploying the patient's electronic record since 2010. In 2014, it was time to abandon the nursing paper documentation file. This represented a user base of 4,000 people to train and accompany during the change process. Design/methodology/approach: We developed the concept of an event as a support for nursing documentation, an empirical process. This project explored a combination of different strategies to improve the quality of care. Research limitations. It is not a research, the concept was new and a theoretical framework was developed. Unfortunately, there was no evaluation undertaken to measure the impact on the users. Practical implications. Today the use of an event helps the nurses to document complex situations and to guarantee the continuity of information. For the future, the question or plan? is to expand the concept to Doctors and other caregivers

Keywords. Business process management, Nursing informatics, Usability, standardized Language, Event, Process modelling

1. Introduction

This poster presents one aspect of the conceptual work whose goal was to describe the nursing activity and roles. The electronic tool acts like a dashboard for a car driver. The concept of EVENT as an episode in which people are involved during an undefined period. These situations might be as different as, daily clinical evaluation, hospital admission, an adverse event, dialysis, surgical intervention, nursing consultation, etc.

To make a simple documentation system that makes the Nursing care complexity visible, prevents a nursing burden being as intuitive as possible

2. Method: Concept of Event and his taxonomy

To document every activity, care, intervention, etc, the work environment must be adapted. In an EHR, this ergonomic facilitation is made possible thanks to the order systems (functionality?) and the related care planning.

The untreated cases are all the situations for which no order is required due to the absence of planning. For example, in the case of a patient admission, a coronary angiography different people are involved, different care and evaluations are made, over a period of time, etc. All these elements are parts of different events.

The taxonomy is in continuous development. The first branch of the taxonomy: Activity contains all the concepts related to the hospital stay. The first domain contains all the movements, the second is dedicated to all types of nursing consultations, the third includes the process related to the management in the operating room, interventional imaging. The fourth is dedicated to administrative activities. The second and third branches integrate the JCAHO taxonomy[2]. These events are dedicated to support the prevention of incidents or the handling of incidents.

We created a documentation model to go out of the focus charting and avoid the model proposed by the Soarian system based on an assessment, nursing diagnoses, interventions and outcomes. ESPOIR [7] for **Event**, Sign & symptoms, Phenomenon, "Objective" (goal), Intervention, Response. This approach is time line oriented. The main change for the nurses was to do bedside documentation in a continuous mode.

3. Results and discussion

The event feature is very effective. Its use has evolved since the beginning of the nursing deployment of the EHR. Today the events implemented are more sophisticated to support nurses in complex clinical situations.

The quality of the documented data has improved beyond that of the old paper age. The different solutions adopted to create the electronic documentation system with the event helped us to face the weakness of the legacy product or paper system? For the future, the question is to expand the functionality to Doctors and other caregivers

Research limitations. It is not a research, even the concept was new, a theoretical framework was developed. Unfortunately, we did not measure the impact on the users.

References

- [1] Lampe, S.S., *Focus charting: streamlining documentation*. Nurs.Manage., 1985.**16**(7): p. 43-46.
- [2] Chang, A., et al., *The JACHO patient safety event taxonomy: a standardized terminology and classification schema for near misses and adverse events*. International Journal for Quality in Health Care, 2005. **17**(2): p. 95-105.
- [3] Yndestad, H. *QUALITY MODELLING USING GENERAL SYSTEM THEORY*. in Fjordkonferansen. 2013. Loen.
- [4] von Berthalanffy, L., *General Systems Theory*. 1968, New-York: Braziller.
- [5] Junger, A., C. Delaney, and A. Berthou, *Modeling, the Essential Step to Consolidate and Integrate a National NMDs*, in *Medinfo04*. 2004, Medinfo: San Francisco.
- [6] Junger, A. and A. Berthou, *NURSINGdata: un modèle de système d'information pour les soins infirmiers*. Hôpital Suisse, 1999(5): p. 25-28.
- [7] Junger, A., H.B. Levi, and G. Falquet, *ESPOIR: from narrative to data oriented documentation*, in *11th International Congress on Nursing Informatics*. 2012, AMIA: Montréal.

Experience and Expertise of Teledermatologists with Teledermoscopy: A Systematic Review to Guide Future Practice

Femke VAN SINDEREN^{ab1*}, Esmée TENSEN^{ab*}, Job VAN DER HEIJDEN^a, Leonard WITKAMP^a, Monique JASPERS^b, and Linda PEUTE^b

^aKsyos Health Management Research, Amstelveen, Netherlands

^bAmsterdam UMC, University of Amsterdam, Department of Medical Informatics, Amsterdam Public Health Research Institute, Meibergdreef 9, Amsterdam, Netherlands

* First and second author contributed equally

1. Introduction

Dermatologists' experience with teledermoscopy (TDsc) and accuracy of teleconsultations seem to be related [1]. If so, the diagnostic accuracy of TDsc could be increased by letting more experienced teledermatologists (TDs) assess the teleconsultations [1]. Expertise with assessing dermoscopic pictures is deemed to be more important than image quality in assessing those TDsc consultations accurately [2]. Our aim is to provide an overview on reported characteristics in literature of TDs' expertise in TDsc.

2. Methods

A systematic literature search was conducted using PubMed (MEDLINE), EMBASE (OVID), and Cochrane Skin Review Group with publications until July 2018. Two reviewers (ET and FS) independently evaluated the eligibility of the full original papers focusing on the detection of skin cancer with TDsc using dermoscopic images. The data was extracted according to a pre-defined data extraction sheet. Based on a grounded theory approach, two main themes were determined for further analysis.

3. Results

47 Articles met inclusion criteria, of which 29 articles reported on years of experience and expertise of the TDs assessing the TDsc consultations (Table 1). The grounded theory approach provided insight on expertise with dermoscopy and pigmented lesions, and experience with TDsc. In five of the 24 studies, TDsc consultations were not assessed by dermatologists, e.g. physicians. Six out of 29 studies mentioned that the TDs were board-certified dermatologists without specific characteristics mentioned of

¹ Corresponding author; Femke van Sinderen. Email: f.vansinderen@amsterdamumc.nl.

the obtained certification. Eight out of sixteen studies expressed explicitly the number of years of experience with TDsc, ranging from two until fifteen years. Seven studies rated the experience by general statements, like ‘low level of experience’.

Table 1. Results of the data extraction concerning the teledermatologists’ expertise and experience

Reported:	Expertise (%)	Number of years of experience (%)
Yes	24 (83%)	16 (55%)
No	5 (17%)	13 (45%)
Total	29 (100%)	29 (100%)

4. Discussion

Data was extracted on the experience and expertise of TDs in assessing TDsc consultations. For a majority of the studies, it is unknown what the expertise in dermoscopy and pigmented lesions is, nor the education in (tele)dermatology. Unclearness in assessor expertise indicate to ambiguity in who should assess the TDsc consultations in order to obtain accurate diagnostic outcomes. Future work should focus on guidelines providing insight into the needed experience to be considered an ‘experienced’ TD. Kittler et al. [5] reported an improved diagnostic accuracy for melanoma for experienced TDs suggesting the need of experience and training to perform TDsc. Training for GPs in TDsc has been proposed [6]. Secondly, future work needs to define clear experience scales and develop training guidelines in TDsc, both in studies and in current clinical practices. We will continue this study by focusing on the relation of TDs’ expertise and experience in TDsc with diagnostic outcomes.

5. Conclusion

We conclude that a minority of studies reported on the TDs’ experience and expertise in diagnostic accuracy studies. It is likely that this will influence the diagnostic accuracy.

References

[1] K. Kravets, O. Vasylenko, Z. Dranyk, and O. Bogomolets, Store-and-forward teledermatology for the most common skin neoplasms in Ukraine, *Acta dermatovenerologica Alpina*, **27** (2018), 79–83.

[2] K. J. Lee and H. P. Soyer, Future Developments in Teledermoscopy and Total Body Photography, *Int. J. Dermatology Venereol.* **2** (2019), 15–18.

[3] M. Pai et al., Systematic reviews and meta-analyses: an illustrated, step-by-step guide, *Natl. Med. J. India*, **17** (2004), 86–95.

[4] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and PRISMA Group, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.,*Ann.Intern.Med.* **151** (2009),264–9.

[5] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, Diagnostic accuracy of dermoscopy, *Lancet. Oncol.*, **3** (2002), 159–65.

[6] O. T. Jones, L. C. Jurascheck, M. Utukuri, M. M. Pannebakker, J. Emery, and F. M. Walter, Dermoscopy use in UK primary care: a survey of GPs with a special interest in dermatology, *J. Eur. Acad. Dermatology Venereol.*, **33** (2019), 1706–1712.

Experience Story: How Do We Re-Implement What Has Been Implemented?

Melanie CASSARINO^{1a}; Camila GALVAN^a; Juan DESCALZO^a; Elio JEREZ^a;
María SMITH^a; Daniel LUNA^a

a Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract. Since 2017, the Hospital Italiano de Buenos Aires, has a “Workstation on Wheels” project where nurses can access to a mobile application in order to register the drug’s administration, vital signs and complete an early warning assessment scale of the hemodynamic state of the patient. Although the overall objective was to achieve at least 95% drug registration through this system, their use did not remain stable over time. Therefore, it was necessary to create an interdisciplinary team to make a diagnosis of the project situation and reasons for the low use rate. In this process, a re-implementation of the barcoding administration system was carried out, focusing on the nursing staff maintaining the use of the system over time. The aim of this paper is to describe the experience and lessons learned in the process of re-implementing the drug barcoding system at the patient’s bedside.

Keywords. BCMA, implementation, nurses, Workstation on Wheels

1. Introduction

The Hospital Italiano de Buenos Aires (HIBA) is an institution accredited by the Joint Commission International. In 2015 the implementation of the correct drug identification system was carried out through the technique of “barcoding” or reading with a QR code scanner on desktop devices [1]. Then, in 2017, mobile devices (tablets) were incorporated into “Workstation on Wheels” (WOWs) structures with transferable tray [2]. The mobile application is developed with IONIC as source code. This implementation was carried out in various hospitalization sectors, both pediatric and adult, including intensive care and general sectors. Although the overall objective was to achieve at least 95% drug registration through this system, usage did not remain stable over time. Therefore, it was necessary to create an interdisciplinary team to make a situation diagnosis. In this process, a re-implementation of the barcoding administration system (BCMA) was carried out, focusing on the nursing staff maintaining high system use over time. The aim of this paper is to describe the experience and lessons learned in the process of re-implementing the drug barcoding system at the patient’s bedside.

2. Methods

At first, the different areas involved in the BCMA project belonging to the

¹ Corresponding author E-mail: melanie.cassarino@hospitalitaliano.org.ar

Department of Health Informatics (DIS) of the HIBA were convened. Forming a transdisciplinary team of 2 informatic nurses, an informatic physician, two computer technicians, a network technician, a programmer and a designer. Tours around the sectors were organized to inquiry nurses about the reasons behind low system usage.

3. Results

During this process, several strategies were used. On one hand, those oriented to organizational change management and training of professionals, and on the other, strategies related to the adaptation, support and operation of technological equipment.

Regarding change management and training: Incorporation of nursing supervisors to dynamics work, group and individual training, tracking users, training material development, communication lines, help desk, communication and dissemination.

Related to adaptation, support and operation of IT equipment: A technical team consisting of specialists in mobile devices and WiFi networks was put together. They reviewed each HIBA inpatient sector to detect failures. Changes were made to the network, to improve broadcast and connectivity. Software improvements were made together with the development team. Applications to restrict general use of the devices were used, allowing only the BCMA application to run.

4. Conclusions

The management of organizational change is the basis for a large-scale implementation that can modify staff workflow [3]. Communication improvement between parties involved in the project showed great changes at the organizational level. It enabled exchange spaces, allowing them to work more fluently. Empowering both nurses and technicians to form a single work team. It generated a feeling of belonging on both sides. It was very important to generate a communication channel in order to provide users with a space to report failures or malfunctions. It was also important for this channel to allow the tracking of tickets, generating improvements in response times and more durable solutions. With active listening, opportunities for improvement can be detected, such as the potential incorporation of functionalities already implemented in the desktop EHR to the bedside workflow, such as risk assessment [4].

References

- [1] Maydana T, Giraldo L, González Z, Schachner B, Mayan J, Luna D, et al. BarCode Medication Administration in ICU: Learning from Our Nurses. *Stud Health Technol Inform.* 2017;245:1095–8.
- [2] Giraldo L, Schachner B, Luna D, Benítez S. Exploring Nurses' Perceptions and Expectations Toward a BCMA Implementation Using a Mobile App and Workstations as a Change Management Strategy. *Stud Health Technol Inform.* 2018;250:134–8.
- [3] Slotnisky D. Transformación digital: cómo las empresas y los profesionales deben adaptarse a esta revolución. Digital House. Coding School.; 2016. 184 p.
- [4] Watzlawick P, Bavelas JB, Jackson DD. Teoría de la comunicación humana: interacciones, patologías y paradojas. Editorial y Librería Herder; 1993. 260 p.

Experiences of National Documentation in Electronic Health Records: The Study Among Occupational Health Nurses

Sari NISSINEN^{a,1}

^a*Finnish Institute of Occupational Health*

Abstract. The Finnish National Patient Data Repository requires all health care professionals to document patient data with a uniform structure. The aim of the study was to gather the experiences of Occupational Health nurses for the usefulness and the ease of use of national patient data structures in electronic health records. The study was carried out in summer 2017. 177 OHNs participated in the study. The respondents felt that the use of the national patient record structures improve the quality of work but doesn't enable to accomplish tasks more quickly. Without patient data and the fluent health information exchange, the patient treatment pathways do not operate properly. The development of patient data structures in EHRs have to be continued.

Keywords. Documentation, Occupational Health Nursing, Occupational Health Care, Electronic Health Records

1. Introduction

The purpose of the eHealth and eSocial Strategy 2020 by the Finnish Ministry of Social Affairs and Health is to provide health care professionals with access to up-to-date patient data and information systems that support their work. In Finland, almost all health care professionals use electronic health records (EHRs). The legislation requires that health care providers save patient data from their own data systems to the national Patient Data Repository (PDR). This requires all health care professionals to document the patient data with a national uniform structure. [1-3] Also the fluent health information exchange among the health care professionals should be done through the national PDR. Thus, the documentation of the patient data has to be executed structurally and consistently. When the data is structured and uniform, its reuse in the annual surveys monitoring health among citizens, and comparison of patient data is also possible. [4, 5] The aim of the study was to gather the experiences of Occupational Health Nurses (OHNs) for the usefulness and the ease of use of national patient data structures in EHRs.

2. Methods

The study was carried out in summer 2017. We used an electronic questionnaire for the data collection. The survey questionnaire was sent to 760 OHNs who were working with pre-selected Occupational Health Services providers. We analyzed the data by using SPSS Statistics 24. The questionnaire consisted of multiple-choice questions

¹ Corresponding Author, Head of Development, Finnish Institute of Occupational Health, Topeliuksenkatu 41 b, 00250 Helsinki, Finland; E-mail: sari.nissinen@ttl.fi.

about usefulness and ease of use of patient record structures. The technology acceptance model (TAM) was used to explain the participants' experiences of perceived usefulness and the ease of use. [6] According to the TAM model, before the use of technology, there is an intention that affects the perceived usefulness (how much the system would enhance job performance) and the perceived ease of use (how easy it is to use the system). The easier a system is to use, the more useful it is perceived to be.

3. Results

177 OHNs participated in the study. Most participants (73%) documented patient data in the patient record during the patient contact, and the rest (27%) after the patient contact. The respondents felt that the use of the national patient record structures improve the quality of work (67%) but doesn't enable to accomplish tasks more quickly (36%). They felt that learning to document with the national patient record structures is easy (81%). It is also easy to remember how to perform documentation tasks using these structures (65%). However, a quite many of the respondents (43%) felt that the documentation with the national patient record structures is rigid and inflexible.

4. Discussion

The patient data is important in health care. Without patient data and the fluent health information exchange, the patient treatment pathways do not operate properly. [4] OHNs play an important role as producers of patient data. In Finland, they make the major of the health check-ups and various health counselling visits. The perceived usefulness and ease of use of technology were identified by the OHNs as important characteristics of the national patient data structure. It is important that national data structures will reduce the burden of OHNs' documentation, rather than increase it. The development of patient data structures in EHRs have to be continued.

5. Conclusions

This study produces new information about national documentation structure that can be used in the development of EHR systems in OHS, and to better serve information exchange among OHNs and the other healthcare professionals. It also enables work ability data to be exchanged in the correct format and in the correct place effectively and reliably.

References

- [1] Act on the Electronic Processing of Client Data in Healthcare and Social Welfare 159/2007.
- [2] Hyppönen H, Hämäläinen P, Reponen J, eds. 2015. E-health and E-welfare of Finland. Checkpoint 2015. Helsinki, Finland: National Institute for Health and Welfare.
- [3] Lehto M, Neittaanmäki P. 2017. Finnish Health Data Environment. Publications of Faculty of Information Technology 35/2017. Jyväskylä, Finland: University of Jyväskylä.
- [4] Ministry of Social Affairs and Health. 2016. Digitalization to Support Health and Wellbeing. Ministry of Social Affairs and Health's Digitalization Guidelines 2025. Publications 2016:5. (Only in Finnish)
- [5] Rosenbloom ST, Stead WW, Denny JC et al. 2010. Generating clinical notes for electronic health record systems. *Appl Clin Inform* 2010;1:232–243.
- [6] Davis F. 1989. Perceived usefulness, perceived easy of use, and user acceptance of information technology. *MIS Q* 1989;13:319–340.

Exploratory Analysis of Animal Bites Events in the City of Buenos Aires Using Data from Electronic Health Records

Gastón Nicolás QUINTANA^{1a} and Santiago ESTEBAN^a

^aHealth Information System Office, Ministry of Health, Buenos Aires City, Argentina

1. Introduction

Rabies is a viral disease transmitted by rabid animal bites or injuries. It is usually preventable with proper wound care, post-exposure prophylaxis and animal zoonoses programs [1]. In order to face this issue, it is important for the health care system to track animal bites and other animal related injuries. Animal bite information is usually registered in free-text form, within the electronic health record implemented in the public health system of the city of Buenos Aires. Thus, we aimed to develop an NLP pipeline that allowed us to extract animal-bites related information from clinical notes and represent it in a dashboard.

Rabies is estimated to cause 59 000 human deaths annually in over 150 countries. Due to widespread underreporting and uncertain estimates, it is likely that this number is a gross underestimate of the true burden of disease. 99% of rabies cases are dog-mediated and the burden of disease is disproportionately borne by rural poor populations, with approximately half of cases attributable to children under 15 [2].

2. Methods

A regular expression natural language processing pipeline was developed in order to extract the type of animal that caused the injury, the type of lesion it caused, which body part was compromised, geographic location of the event and whether the patient received rabies vaccine treatment. This information was represented in a dashboard for monitoring and evaluation.

3. Results

From January 2017 until December 2019, a total of 18676 animal bite events were registered. Dogs were responsible for nearly 50% of the events. These injuries occurred mainly in the south of the city, mostly in areas characterized by precarious housing. On the other hand, bat bites occurred close to the city center, an area with higher density of apartment buildings. The number of events tends to be regular over the year but winter

¹ Gastón Nicolás Quintana. gastonquintana@buenosaires.gob.ar.

months have the lowest number of cases. Overall, male and female are equally affected but the cases are concentrated in the 0 - 15 age group, involving mostly dog bites.

4. Discussion

Free-text encoding is a very important source of information and physicians prefer it mainly because it is flexible. The development of this pipeline made evident asymmetries not only in geographic case distribution but also in age ranges distribution. This could be useful to create new strategies or improve existing ones.

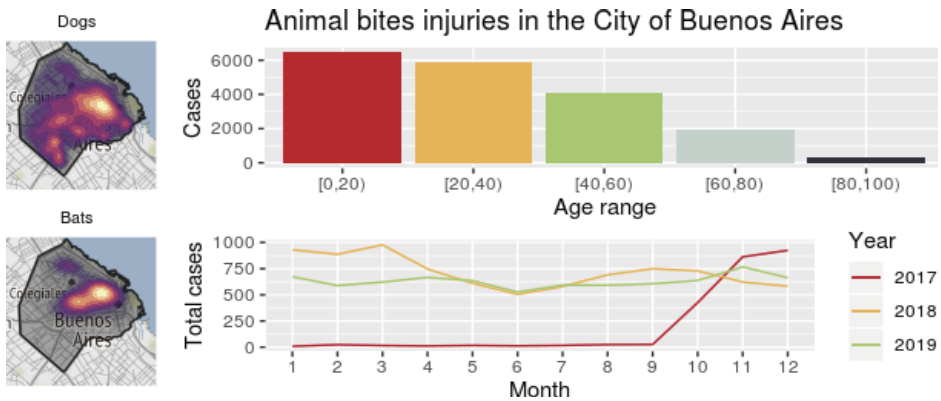


Figure 1. Plots included in the dashboard

5. Conclusion

We described how we managed to extract clinical information stored in the electronic health records using NLP algorithms. As health information is crucial we believe this is a very useful tool to help the ministry and other areas to design public health strategies.

References

[1] Demaria, A. Clinical manifestations and diagnosis of rabies. In: UpToDate, Post, TW (Ed), UpToDate, Waltham, MA, 2020.

[2] World Health Organization. Rabies – Epidemiology and burden of the disease. Available at <https://www.who.int/rabies/epidemiology/en/>. Accessed December 2019.

Exploratory Analysis of Consultation Patterns Prior to the Diagnosis of Depression

Sofia SCIARRETA^{a1} and Santiago ESTEBAN^a

^a*Family and Community Medicine Division, Hospital Italiano, Buenos Aires, Argentina*

Abstract. Depression is frequently underdiagnosed or the diagnosis occurs in late stages, mainly because it usually presents as a paucisymptomatic disorder. Electronic health records (EHRs) are a very useful tool to capture the longitudinal medical history of the patient which may help in diagnosing these patients earlier. We aimed to understand how these patients' consultation patterns are represented in the EHR up until depression diagnosis.

Keywords. Depression, Electronic health records, Depressive Symptoms

1. Introduction

Depression is the most frequent psychiatric medical disorder observed in primary care [1]. The main problem that depression presents is the fact that 50% of the cases are not diagnosed [2][3]. Underdiagnosis is very frequent due to the atypical and varied clinical presentation of depression[3]. Usually, looking at the patients' history and consultation pattern facilitates arriving at a diagnosis. EHRs help to collect and review patient data longitudinally. The objective is to describe the consultation patterns of patients before they are diagnosed with depression at Hospital Italiano de Buenos Aires using data from EHRs.

2. Methods

We manually reviewed a random sample of 600 adult patients over 18 that had at least one depression related code registered in their EHR between 2010 and 2016. 278 had a first diagnosis of depression. We obtained several variables related to healthcare system use. Finally, we explored patient clusters using non-supervised algorithms like k-means and hierarchical clustering. A six cluster solution was found to be the best fit. Clusters were then manually inspected and qualitatively described.

¹ Corresponding author. E-mail: sofia.sciarreta@hospitalitaliano.org.ar

3. Results

Table 1 shows the patients’ characteristics and describes each cluster. Although we statistically obtained a 6 cluster solution, we combined them into 4 according to their main characteristics (Cluster 1, clusters 2 + 3, cluster 4, clusters 5 + 6). Cluster 1: We believe these patients already had a diagnosis of depression, due to the fact that they presented a large number of consultations within a short time of affiliation. In clusters 2+3, there were a large number of consultations to specialties, general medicine, emergency room visits and hospital admissions. They present a pattern of consultations compatible with patients who are underdiagnosed and thus keep seeking care. Cluster 4 and 5+6 comprise most of the sample. Patients in these clusters had most of their visits with their general physician, with different follow up times. They most likely represent patients who take advantage of a primary care centered health system.

Table 1. Patients’ characteristics and description of clusters

Cluster no.	1	2 + 3		4	5 + 6	
n	6	16	4	53	149	50
Age at diagnosis (median [IQR])	72 [58.75, 76.25]	68.50 [62.2, 80]	64.5 [37, 87.5]	65.00 [54, 72]	67.00 [51, 79]	64.5 [46.5, 74]
Women (%)	6 (100.0)	11 (68.8)	4 (100.0)	48 (90.6)	149 (100.0)	0 (0.0)
Seniority	0.30 [0.26, 0.61]	7.47 [1.90, 14.10]	7.95 [4.90, 10.34]	1.57 [0.75, 2.29]	8.27 [5.85, 12.51]	8.89 [6.50, 14.90]
Emergency room visits	0.00 [0.00, 0.00]	0.89 [0.40, 1.30]	1.50 [0.90, 2.48]	0.00 [0.00, 0.82]	0.22 [0.00, 0.60]	0.60 [0.20, 0.98]
Hospital admission	0.00 [0.00, 0.00]	0.00 [0.00, 0.39]	0.50 [0.39, 0.85]	0.00 [0.00, 0.00]	0.00 [0.00, 0.20]	0.00 [0.00, 0.00]
Follow-up years	0.30 [0.26, 0.61]	5.00 [1.90, 5.00]	5.00 [4.46, 5.00]	1.57 [0.75, 2.29]	5.00 [5.00, 5.00]	5.00 [5.00, 5.00]
No. visits per year	203.15 [145.7, 408.7]	62.10 [47.8, 91.2]	42 [37, 47.7]	38.3 [24.3, 64.2]	13.2 [5.4, 22.4]	6.9 [3.5, 16.7]

4. Conclusions

Longitudinal data in the EHR was useful for differentiating clusters of patients with a recent diagnosis of depression.

References

[1] Thompson C, Ostler K, Peveler RC, Baker N, Kinmonth A-L, Dimensional perspective on the recognition of depressive symptoms in primary care. *Br J Psychiatry* (2001), 317-323.

[2] Henkel V, Mergl R, Kohlen R, Maier W, Möller H-J, Hegerl U. Identifying depression in primary care: a comparison of different methods in a prospective cohort study. *BMJ Br Med J.* (2003), 200-201.

[3] Tylee A, Gandhi P. The Importance of Somatic Symptoms in Depression in Primary Care. *Prim Care Companion J Clin Psychiatry* (2005), 167 -176.

Exploratory Analysis of Exercise Mentions in Clinical Notes

Nahuel ORCAIZAGUIRRE^{a,1} Santiago ESTEBAN^a

^a*Family and Community Medicine Division, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina*

Abstract. Physical activity has been associated with many positive health outcomes. In this study we explored clinical notes to better understand how physicians record information on this topic.

Keywords. physical activity, electronic health record

1. Introduction

Physical exercise is defined as any body movement that increases energy consumption above the baseline value [1]. It has been associated with improvements in mood [2], quality of sleep, cognitive level, bone density, obesity and type 2 diabetes prevention[3]. Given this, it has become a first line treatment strategy in primary care for many chronic diseases, like cardiovascular disease. For this reason, we set out to analyze clinical notes at Hospital Italiano de Buenos Aires to assess mentions of physical activity recorded.

2. Methods

As a first step we developed a keyword based filter to detect different types of exercise mentions. We evaluated the filter's performance on a random sample of 2047 clinical notes that we manually reviewed and annotated (presence or absence of physical activity mentions). Afterwards, we applied the filter and obtained a random sample of 750 clinical notes which we manually reviewed to identify which types of exercise patients report doing, which ones physicians prescribe and why they prescribe them to their patients.

3. Results

After manual review, we detected that 157 out of 2047 clinical notes had physical activity mentions (7.7%, 95% Confidence interval (CI) 6.59 - 8.9%). The keyword

¹ Dr. Nahuel Orcaizaguirre, Hospital Italiano Buenos Aires, Pres.Tte.Gral. Juan Domingo Peron 4190, Buenos Aires, Argentina. Email: nahuel.orcaizaguirre@hospitalitaliano.org.ar.

filter yielded a sensitivity of 0.977 (95% CI 0.910 - 0.996), positive predictive value of 0.840 (95% CI 0.750 - 0.903) and F1 score of 0.903 (95% CI 0.857 - 0.942).

After reviewing the 750 clinical notes we observed that physicians mostly (68%) prescribe physical activity in a very generic way (recommendations of 'starting physical activity' or 'joining a gym'). Less frequently (32%), they specify a certain type of activity, usually walking (23%). In most cases there was not enough information as to why a particular type of physical activity had been prescribed. Regarding the reasons why physical activity is being prescribed, the main ones were cardiovascular prevention (includes the treatment of risk factors such as hypertension or diabetes, 42%), treating obesity (25%) and improving general body aches (11%).

4. Discussion

Our results indicate that physical activity is seldom registered in the clinical notes and when it is, there is a lack of information on what it is being prescribed for. We believe that non-pharmaceutical treatments should be registered as rigorously as pharmaceutical treatments, since they are part of many treatment plans. Not having this information available hinders the decision-making process in future consultations.

Furthermore, we see that there is not much variety in the activities suggested by medical professionals. It would be interesting to analyze in the future if this can be related to limited knowledge about the different options available.

5. Conclusion

In our study, physical activity mentions were present in a small proportion of clinical notes in the outpatient setting. Furthermore, information regarding the cause for prescription was scarce. We emphasize the importance of recording non-pharmaceutical treatments rigorously.

References

- [1] Caspersen CJ, Powell KE, Crhistenson GM, Physical activity, exercise, and physical fitness: definitions and distinctions for heal-related research. *Publi Health Rep.* **100(2)** (1985), 126-131
- [2] King AC, Taylor CB, Haskell WL, DeBusk RF. Influence of regular aerobic exercise on psychological health: A randomized, controlled trial of healthy middle-aged adults. *Heal Psychol.* **8(3)** (1989), 305-324.
- [3] Physical Activity Guidelines Advisory Committee. (2008). Physical activity guidelines advisory committee report, 2008. *Washington, DC: US Department of Health and Human Services, 2008.*

Feature Set for a Prediction Model of Diabetic Kidney Disease Progression

Masaki ONO^{a,1}, Takayuki KATSUKI^a Masaki MAKINO^b Kyoichi HAIDA^c
Atsushi SUZUKI^b and Reitaro TOKUMASU^a

^aIBM Research - Tokyo

^bThe Dai-ichi Life Insurance Company, Limited

^cFujita Health University

Abstract. In this paper, we propose feature extraction method for prediction model for at the early stage of diabetic kidney disease (DKD) progression. DKD needs continuous treatment; however, a hospital visit interval of a patient at the early stage of DKD is normally from one month to three months, and this is not a short time period. Therefore it makes difficult to apply sophisticated approaches such as using convolutional neural networks because of the data limitation. The propose method uses with hierarchical clustering that can estimate a suitable interval for grouping inputted sequences. We evaluate the proposed method with a real-EMR dataset that consists of 30,810 patient records and conclude that the proposed method outperforms the baseline methods derived from related work.

Keywords. disease risk prediction model, diabetic kidney disease

1. Introduction

We constructed a risk prediction model for diabetic kidney disease (DKD) that receives the medical events of a patient in the past 180 days and predicts whether the patient will stay at the 1st stage after 180 days. A patient at the early stage of DKD needs continuous treatment; however, a hospital visit interval is from one month to three months, and this is not a short period. We focused on this issue, and the proposed method derives time series variation with hierarchical clustering that can estimate a suitable interval for grouping inputted sequences.

The number of studies on risk prediction models has increased and most of the studies applied logistic regression or cox regression to the latest medical events. Some studies focuses on patients that frequently visit a hospital, so records with many medical events in a certain period of time are available [1][2] and use CNN to distill a time series variation of a given patient.

The proposed method transforms a value sequence into groups on the basis of the result of the hierarchical clustering for a corresponding timestamp sequence. We use the hierarchical clustering method that has the constraint not on the number of the clusters but on the distances among the clusters. Then, we derive the time series variation from

¹E-mail: moonoo@jp.ibm.com

Table 1. The AUC in the prediction for the stage of DKD of the patient after 180 days. “LR” stands for the logistic regression.

ID	Model	AUC	ID	Model	AUC
1	LR without time series variation	0.625	4	CNN #1 (interval 30 days)	0.632
2	LR with time series variation	0.722	5	CNN #2 (interval 14 days)	0.660
3	CNN #1 (interval 14 days)	0.666	6	CNN #2 (interval 30 days)	0.635

the group. For example, we calculate the averaged value and the standard deviation from the averaged cluster’s values. The key factor is the hierarchical clustering that can estimate a suitable interval for grouping inputted sequences. We use logistic regression as a statistical model for the prediction because logistic regression can provide interpretable results. Interpretation is one of the important factors in the field of medical informatics.

2. Results and Discussion

We evaluate the proposed method with a real-EMR dataset from a Japanese hospital. We collected the medical history of a patient at the 1st stage. The number of patients is 30,810, where half of the patients remain at the 1st stage after 180 days. We focus on 12 kinds of laboratory tests as feature sets on, for example, uric protein and eGFR.

We conducted a 5-fold cross-validation test in the experiment and measured the AUC, which is the standard evaluation metric for a classification problem. And we used the baselines, the first of which is a logistic regression that handles the latest values of the feature sets and does not handle the time series variations that the proposed method creates. The other baselines are two kinds of CNNs that consist of 5 layers and handle feature matrices created in a previous work [1] [2]. We created two kinds of feature matrices created with different the interval for CNNs, 14 days and 30 days.

Table 1 shows the experimental results. The best performance is 0.722, which was achieved by the logistic regression with the time series variation that the proposed method creates. The experimental results point to the following conclusion: (1) understanding the time series variation improves the prediction performance of the risk prediction model of diabetic nephropathy and (2) the proposed method outperforms the baselines. The first conclusion is reached on the basis of the comparison between the ID 1 result and the others; the AUC of ID 1 is lowest and is derived from the logistic regression without the time series variations. The second conclusion is reached on the basis of the comparison between the ID 2 result and the others. The ineffectiveness of CNN for this task is caused by the missing value in the feature matrix. A low frequency of hospital visits causes the missing value in the feature matrix for CNN.

References

[1] Y. Cheng, F. Wang, P. Zhang and J. Hu, Risk Prediction with Electronic Health Records: A Deep Learning Approach, in: *SDM*, 2016.
[2] Z. Huang, W. Dong, H. Duan and J. Liu, A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records, *IEEE Transactions on Biomedical Engineering* **PP**(99) (2017), 1–1. doi:10.1109/TBME.2017.2731158.

FHIR Driven Self-Management Support System for Diabetes

Dietmar GLACHS^{a1}, Tuncay NAMLI, Oliver JUNG^a, Felix STROHMEIER^a, Manuela PLOESSNIG^a and Gustavo RODRIGUEZ^c

^aSalzburg Research Forschungsgesellschaft mbH, Salzburg, Austria

^bSoftware Research & Development & Consultancy Corp., Turkey

^ciHealthLabs Europe SAS, France

Abstract. The number of people with diabetes is increasing in every European country and as with all chronic diseases, coping with diabetes is a long-term process. Self-Management and supporting behaviour change are aspects when dealing with diabetes. The POWER2DM² system combines treatment planning and self-management activities by providing interventions to change a patient's lifestyle towards a healthier, diabetes-appropriate life. FHIR³ allows for data exchange.

Keywords. decision support systems, self-management, behaviour change

1. Introduction

In this paper, we introduce the Self-Management Support System (SMSS) for Type 1 and Type 2 diabetes patients based on the Personal Data Store (PDS), a fully FHIR compliant data repository providing electronic health records. Empowering patients [1] by integrating them to a greater extent with their own healthcare process is the overall target. Disease management becomes an integrated part of their life.

2. Method

Self-Management Support is implemented as a guided iterative workflow cycle, typically on a weekly basis. Treatment plans are usually generic and span longer periods of time. Patients are requested to adopt the practitioner's treatment plans into personal goals and to plan activities he or she wants to accomplish within a shorter period of time. Increasing the patient's adherence to the personal action plan is achieved by providing interventions based on Behaviour Change Techniques [2]. Basically, the workflow consists of four main steps as shown in Figure 1. First the patient specifies self-management *Goals* and relates them to the practitioners' treatment plan.

Secondly, *Actions* like tasks and activities are planned with a Calendar. Again, tasks and activities are related to self-management and treatment goals. In the *Monitoring*

¹ Corresponding Author: Dietmar Glachs, Salzburg Research Forschungsgesellschaft mbH, Jakob-Haringer-Straße 5/3, 5020 Salzburg, Austria, E-Mail: dietmar.glachs@salzburgresearch.at

² Funded by the EU Horizon 2020 research and innovation programme under grant agreement No 689444

³ "Fast Healthcare Interoperability Resources" (<http://hl7.org/fhir>).

phase, blood glucose, blood pressure and body weight are recorded with devices. Activities such as medication intake, physical exercises, meals are recorded manually. It is also possible to keep a diary for problems, mood, sleep and stress. Finally, with the Periodic Review, the recorded data is *evaluated* and performance *feedback* is provided. For each of the review categories, positive reinforcement messages [3] or recommendations for the next action plan cycle are given. In [4], the intervention mechanisms are explained in detail.



Figure 1. Patient Empowerment Workflow

Data is maintained FHIR compliant with the PDS while ensuring interoperability. 39 FHIR “Profiles” and FHIR “Extensions” customize basic FHIR resources and contribute SMSS specific data elements [5]. A detailed architecture of the POWER2DM system is given in [6]. The FHIR profiles are available at Simplifier.net⁴.

3. Conclusion

The Self-Management Support System supporting behaviour change is currently evaluated in pilot applications in a production environment with 230 patients (115 type-1, 115 type-2) in the Netherlands and in Spain. The trial evaluates the acceptance rate and effectiveness of the presented interventions. Results are expected by the end of the pilot applications. A test system is also available and adjustable to other chronic diseases.

References

- [1] Castro, E., Regenmortel, T., Vanhaecht, K., Sermeus, W., Hecke, A., Patient empowerment, patient participation and patient-centeredness in hospital care: A concept analysis based on a literature review, *Patient Education and Counseling* **99**(12) (2016), 1923-1939.
- [2] Abraham, Ch., Michie, S., A Taxonomy of Behavior Change Techniques Used in Interventions by the American Psychological Association, *Health Psychology* **27**(3) (2008), 379–387.
- [3] Elder, J., Ayala, G., Harris, S., Theories and intervention approaches to health-behavior change in primary care, *American Journal of Preventive Medicine* **17**(4) (1999), 275-284.
- [4] Jung, O., Glachs, D., Strohmeier, F., Mulrenin, R., Huisman S., Smith, I., van Keulen, H., Sont, J., Ploessnig, M. (2019): Empowering Diabetes Patients with Interventions Based on Behaviour Change Techniques In: dHealth 2019 - From eHealth to dHealth. Proceedings of the 13th Health Informatics Meets Digital Health Conference. S. 154-161.
- [5] Namli, T., Köse, O., POWER2DM D4.2 Personal Data Store Service Implementation, <http://power2dm.eu/wp-content/uploads/POWER2DM-D4.2.pdf>, last access: 23.9.2019.
- [6] Namli, T., POWER2DM D1.3 Conceptual Design of the POWER2DM Architecture, <http://www.power2dm.eu/wp-content/uploads/Power2DM-D1.3-1.pdf>, last access: 23.9.2019.

⁴ POWER2DM project account: <https://simplifier.net/POWER2DM>

GDPR Compliant Blockchain and Distributed Ledger Technologies in the Health Sector

Anton HASSELGREN ^{a,1}, and Katina KRALEVSKA ^b, Danilo GLIGOROSKI ^b and Arild FAXVAAG ^a

^a*Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, NTNU- Norwegian University of Science and Technology, Trondheim, Norway*

^b*Department of Information Security and Communication Technologies, Faculty of Information Technology and Electrical Engineering, NTNU-Norwegian University of Science and Technology Trondheim, Norway*

Keywords. Blockchain, Distributed Ledger Technology, GDPR, Health Data, Protected Health Information

1. Introduction

In 2018, The European Union instituted the General Data Protection Regulation (GDPR), which regulates the collection, processing, and securing of personal data, including protected health information (PHI). As detailed by the European Union Blockchain Observatory and Forum [1], in principle, there are no contradictions between the goals of GDPR and Distributed Ledger Technologies (DLT). However, there seems to be at least three areas in which GDPR still does not offer enough clarity about how real-world DLT applications for the health sector should be developed. These areas include (1) accountability and roles (eg, how to identify a data controller in a public DLT), (2) anonymization of personal data (eg, what techniques are sufficient to anonymize personal data to the point where the resulting output can potentially be stored in a DLT), and (3) GDPR rights conflicts (eg, how to rectify or remove personal data that are recorded in a DLT that is immutable by nature, or who is responsible for requesting and managing the “freely, specific, informed, and unambiguous” consent from a data subject, especially if the data controller is not specified) [2]. With regards to anonymization of personal data, it is clear that GDPR does not apply to anonymized data and that this type of information can be stored on the DLT. However, what qualifies as anonymized is still not clear. The only indication today is that it must be irreversibly impossible to identify an individual through any of the means “reasonably likely to be used” [3]. The objective of this study is to dissect the various designs of DTL and look for GDPR compliance for the different components in established or proposed blockchain/DTL applications that deals with health data.

¹ Corresponding Author, Mellomila 69, 7018 Trondheim, Norway; E-mail: anton.hasselgren@ntnu.no

2. Method

A comprehensive review over the current GDPR official documents and other reports that touch upon the subject, including the report developed by the EU blockchain Observatory and Forum [1]. Furthermore, a summary of peer-reviewed publications under the topic and retrieval of white papers for proposed applications in the health sector. Literature reviews of high quality that summarize the current state-of-art and state-of-knowledge should be explored. Summarize the key aspect of the reviewed literature and the key components of blockchain/DTL will be based on peer-reviewed summaries/literature reviews.

3. Results

Our preliminary results indicate that there are several areas where blockchain/DTL in the health sector may face challenges when it comes to GDPR. More specifically, a public blockchain that contains on-chain data related to an individual's health, faces serious challenges when guaranteeing the anonymization of that data.

4. Discussion

There are clear indications that carefully considerations should be undertaken when designing a blockchain infrastructure that deals with health data, to comply with the current GDPR. As previous research shows, the developer of a blockchain system have several important design decisions to make that affect the privacy, access and accountability of the system. It is of high importance to consider the GDPR and especially the three areas mention in the introduction.

5. Conclusion

How blockchain and DTL comply with current regulations for health data have been suggested to be one main barrier for implementation of this technologies in the health sector. Our preliminary result shows that when certain design choices are made, blockchains and DTL for the health sector may not only comply with the current GDPR but will contribute to reinforce the regulations.

References

- [1] European Union Blockchain Observatory and Forum. Blockchain and the GDPR. 2019. URL: https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf?width=1024&height=800&iframe=true
- [2] Official Journal of the European Union. 2016 Apr 14. Regulation (EU) 2016/679 of the European Parliament and of the council URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679>
- [3] Working Party on the Protection of Individuals with Regard to the Processing of Personal Data. Opinion 05/2014 on Anonymisation Techniques. 2014 Apr 10. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Generating Surveillance Data for Nosocomial Infections from Routine Charting in Intensive Care Units

Achim MICHEL-BACKOFEN^a, Dorothee STEINKE^b, Christian KATZER^c, Florian BRENNCK^c, Andreas BÖNING^d

^a*Department for clinical and administrative data processing, University Hospital of Gießen and Marburg, Gießen*

^b*Institute for hospital hygiene, University Hospital of Gießen and Marburg, Gießen*

^c*Department for anesthesiology, surgical intensive care medicine and pain therapy, University Hospital of Gießen and Marburg, Gießen*

^d*Department for cardio surgery, University Hospital of Gießen and Marburg, Gießen*

Abstract. The University Hospital of Gießen and Marburg (UKGM), Gießen, Germany participates in a country wide nosocomial infections surveillance project on intensive care units (ICU) called KISS. KISS data must be gathered daily and is comprised of detailed clinical data like patient movements, ward utilization, presence of central venous and urinary catheters and types of artificial respiration. This happens for all of the 10 ICU's at UKGM Gießen and proved to be very tedious and time consuming for the institute for hospital hygiene. The goal of this project was the automated generation of the KISS reports from routinely collected data in the patient data management system (PDMS). The results show that this is largely feasible without changing the documentation habits of physicians and nurses.

Keywords. PDMS, hospital hygiene, nosocomial infections, infection surveillance

1. Introduction

The University Hospital of Gießen and Marburg (UKGM), Gießen, Germany participates in a country wide nosocomial infections surveillance project on intensive care units (ICU) called KISS [1]. All participants of the KISS study must provide a set of clinical data items on a daily basis. The items are comprised of counters for patient movements, ward utilization, presence of central venous and urinary catheters and types of artificial respiration. They are collected for all of the 10 ICU's at UKGM Gießen. This is a very tedious and time-consuming process for the members of the institute for hospital hygiene. The goal of this project was to derive automatically all parameters of the KISS report from routinely collected data in the patient data management system (PDMS). This should happen without changing the established documentation habits on the wards.

2. Material and Methods

The University Hospital of Gießen and Marburg (UKGM), Gießen, Germany operates a patient data management system (PDMS) on all intensive care units (ICU's) and intermediate care units (IMC's) since 1999 [2]. The documentation is based on a controlled vocabulary that covers all aspects of intensive care including application of intravenous and urinary catheters, applied drugs and presence and type of artificial respiration. A background application has been developed, that routinely checks the PDMS data for the presence of information regarding the KISS study parameters. This is accomplished by checking the presence of a set of device related concepts from the vocabulary. The analysis is done for the complete course of a patient stay to keep track of application and removal of the different types of catheters. All results are collected in a separate hygiene database. The content of the hygiene database is displayed with an ASP.NET Web-Application. The GUI layout resembles the layout of the KISS data entry form and therefore allows for seamless copy and paste of information.

3. Results

Development started in early 2018. Since October 2018 the system is used for cross checks of the manually collected data which acts as a gold standard. Quality checks and optimizations are currently applied separate for every clinical department. Data shows a sensitivity of 1.0 for detection of admissions and discharges, 0.95 for urinary catheters and 0.9 for central venous catheters and artificial respiration. Specificity for admissions and discharges is 0.8 and 0.9 for urinary catheters, central venous catheters and artificial respiration. Incomplete documentation of device application or removal has been detected as the main reason for failure of the automated process. Additional analysis of the clinical context e.g. balance data for urinary catheters, has proven to be a valuable source of information to cope with this documentation deficits.

4. Discussion

The developed application allows for a great amount of time saving for the members of the institute for hospital hygiene. It removes the need to manually extract data from clinical charts on a daily basis which proves to be an error prone and tedious process.

In future the data extraction system will be enhanced to extract more data which could then be used to act as an early warning system for nosocomial infections.

References

- [1] Gastmeier P, Geffers C, Koch J, Sohr D, Nassauer A, Daschner F, Rüden H. Surveillance nosokomialer Infektionen: Das Krankenhaus-Infektions-Surveillance-System (KISS). *J Lab Med* 1999; 23:173-178
- [2] Michel, A., Benson, M., Junger, A., Sciuk, G., Hempelmann, G., Dudeck, J., Marquardt, K.: Design Principles of a Clinical Information System for Intensive Care Units (ICUData), In: Hasmann, A., Blobel, B., Dudeck, J., Engelbrecht, R., Gell, G., Prokosch, H.-U., Medical Infobahn for Europe, Proceedings of MIE2000 and GMDS2000, Studies in Health Technology and Informatics (Vol. 77), IOS Press, Ohmsha, 2000.

Great Help for Small People – The Development of a Children's Emergency App

Judith BORN^{a,1}, Christina OHLMEIER^a, Jasmina KIS^b, Ralf SICK^c, Jörg OBERFELD^a
and Christian JUHRA^a

^aUniversity Hospital Muenster, Germany

^bHELIOS S. Josefs Hospital Bochum, Germany

^cSt. Johns Ambulance Service, Germany

Abstract. In order to assist parents in providing rapid first aid to children in case of emergency, a mobile emergency application was developed. The children's emergency app works with intuitive graphics and checklists and combines several functions such as tips for preventing accidents, easy-to-understand instructions for first-aid actions and a search function for the next pediatrician or hospital. This app is not intended to be a substitute, but rather an addition to first aid courses.

Keywords. Mobile apps, children's health, first aid, prevention

1. Introduction

Accidents are one of the highest risks to children's health. The German Federal Ministry of Health estimates that in Germany medical care is given to around 1.7 million children who experience accidental injuries each year [1]. An emergency of one's own child is one of the greatest worries for any parent. When health emergencies occur and seconds count, adults must take the right actions to save lives. They need to feel confident within the situation to be able to do so, but in fact many people are afraid to do something wrong in case of an emergency [2]. For this reason the St. Johns Ambulance Service Germany, the BARMER health insurance, the mobile application development company Opwoco and the University Hospital Muenster developed an mobile application, which is intended to help parents, tutors or teachers to initiate necessary steps in emergency situations until paramedics or doctors can intervene.

2. Method

The cornerstone for the app was placed within a project called MARIKA, which was funded by the European Union and the Federal Ministry of Health, Emancipation, Care and Aging of North-Rhine Westphalia. Within this project a survey on user interests and needs in the field of mobile health applications was conducted among 316 patients and accompanying persons at the University Hospital of Muenster [3]. On the basis of these findings and on the basis of the guidelines of the European and German

¹ Corresponding Author: Judith Born, University Hospital of Muenster, Hüfferstraße 73-79, 48149 Münster, E-Mail: judith.born@ukmuenster.de

Resuscitation Councils as well as the experiences of the nationwide "First Aid for Children" courses of the St. Johns, the children's emergency app was developed.

As a database for the location-based search function for pediatric and surgical ambulances, the quality reports of the clinics were used. These are published annually by all clinics in Germany.

3. Results

The children's emergency app meets all data protection requirements of the Federal Data Protection Act in Germany and can be used free of charge on Android and iOS. It works with intuitive graphics and checklists, so it is easy to understand and intuitive to use and combines several important functions:

- search for the next pediatrician, hospital, etc.
- setting up an emergency call
- easy-to-understand instructions for first-aid actions
- tips for preventing childhood accidents
- individually fillable address book for important medical numbers

4. Discussion

The number of downloads (107.688 downloads within two years) and ratings on the download platforms indicate a high user acceptance. The app is refined continuously, so in the beginning of 2019 an updated version was provided, which was extended by the topic "childhood diseases". In addition, the app has been provided with a new feedback function. Since the app is only available in German so far, the translation into different languages will also be considered.

5. Conclusion

The children's emergency app is not a substitute, but a valuable addition to first aid courses on the child. Even with the newly integrated feedback function, the greatest challenge for the developers remains to keep the contact data for all hospitals, paediatricians and pharmacies up to date, as there is a lack of a central and current database for all health care providers in Germany.

References

- [1] Federal Ministry of Health of Germany, Prevention of children's accidents, URL: <http://www.bundesgesundheitsministerium.de/themen/praevention/kindergesundheit/praevention-von-kinderunfaellen.htm> Erste Hilfe und Wiederbelebung! [Accessed: 16 August 2019]
- [2] Forsa, Erste Hilfe und Wiederbelebung, URL: https://www.johanniter.de/fileadmin/user_upload/Dokumente/JUH/BG/Meldungen/Umfrage_Forsa_Ergebnisse_Erste_Hilfe.pdf [Accessed: 7 January 2020]
- [3] J. Mamuzic, Mobile Health als Teil der modernen Medizin Anforderungen aus Sicht der Bevölkerung, Dissertation zur Erlangung des doctor medicinae der Medizinischen Fakultät der Westfälischen Wilhelms-Universität Münster, 2016.

Greek Hospitals Web Accessibility

Andriana MAGDALINO^{a,1} and John MANTAS^a

^aNational and Kapodistrian university of Athens, Greece

Abstract. Even though Internet is an exciting technological tool, it requires innovative design to be accessible to everyone, including people with disabilities. The purpose of the present work is to evaluate the accessibility of the web sites of the hospitals which are under the Greek Ministry of Health and belong in the first Health Region using a free WCAG 2.0 and Section 508 Web Accessibility Check software. An observation of the websites along with the software showed that no website meets the W3C criteria of accessibility.

Keywords. Accessibility, website, disability

1. Introduction

Internet access equality is important because of the impact on peoples' health and on society [3]. At the social level, Internet accessibility and use is associated with increased citizens' trust, interpersonal trust and life satisfaction [4]. People with disabilities seem to be facing a significant digital divide having access to different types of services [1], [9], [11]. People with disability are usually not independent with health information retrieval [2]. More precisely, people with vision impairment tend to use screen readers to browse the Web. For this purpose, web sites elements must be accessible and readable from screen readers [8]. The Web Content Accessibility Guidelines (WCAG) [7] were developed by the World Wide Web Consortium (W3C) [5] and are widely accepted as standards for digital accessibility [10]. The purpose of this article is to evaluate the web accessibility of the public hospitals which are under the Ministry of Health [6] and belong in the first Health Region in Greece.

2. Methods

Twenty five hospitals that belong to the first Health Region were identified from the official website of the Ministry of Health and twenty one of them had website. The free WCAG 2.0 and Section 508 Web Accessibility Check tool [5] was used to automatically scan the websites and evaluate accessibility based on the WCAG 2.0 criteria. Website accessibility issues were distributed in the following categories: Font, Images, Tables, Headings, Links, Colours, Labels, Pop ups, Tags, HTML element missing, Sitemap missing. Hospitals with one or more accessibility issues did not meet the W3C criteria and they were rated as non accessible.

¹ Corresponding Author: Andriana Magdalino, National and Kapodistrian University of Athens, Papadimantopoulou 123, Athens, Greece; E-mail: alfami@nurs.uoa.gr.

3. Results

Twenty one hospital websites were evaluated giving the following results: 29% of the hospital websites had Font size issues. 72% of the hospital websites had images related issues such as images with empty alternate text, images inside an anchor tag without alternate text, animated images and images without alternate text. 48% of the hospital websites had issues Tables. 38% of the hospital websites had issues with missing or empty Headings. All hospitals websites had issues with empty links or missing skip links and links with ambiguous link text. Moreover elements to be using an event handler such as on click to emulate a link were found. 33% of the hospital websites had issues with the colour background. More precisely, an element was found to not have a contrast ratio of at least 4.5:1 between the foreground and background colours and no buttons for background colour alteration and letter size alteration was found. 29% of the hospital websites had a form field found to be missing labels. 76% of the hospital websites had issues with pop up windows. 38% of the hospital websites had issues with the tags. 48% of the hospital websites HTML element was missing. 38% of the hospital websites had no sitemap. No hospital website met all the WCAG 2.0 criteria.

4. Conclusions

The main issues found in the majority of hospitals websites were related to images, links and pop up windows. More specifically missing skip links, untitled tables, and images without alternate text were identified. The hospital websites included in this work do not meet the WCAG 2.0 specifications and therefore cannot be considered accessible. These hospital websites should be updated in order to comply with accessibility standards.

References

- [1] M. Duplaga, Digital divide among people with disabilities: Analysis of data from a nationwide study for determinants of Internet use and activities performed online *PLoS One* **12** (2017).
- [2] M. Thurston, A. Thurston, The accessibility of health information for blind and partially sighted people, *RNIB Scotland* (2010).
- [3] S. Park, J. Humphry, Exclusion by design: intersections of social, digital and data exclusion, *Information, Communication & Society* **22** (2019).
- [4] J. Davis, The Accesibility Divide: The Visually-Impaired and Access to Online News, *47 J. Broad. & Elec. Media* **474** (2003).
- [5] W3C Available at: <https://www.w3.org/WAI/policies/european-union/>
- [6] Ministry of Health, Greece Available at: http://www.lydye.gov.gr/?page_id=70
- [7] WCAG 2.0 and Section 508 Web Accessibility Check. Available at: <https://www.alumnionlineservices.com/scanner/>
- [8] I.V. Ramakrishnan, V. Ashok, S.M. Billah, Non-visual Web Browsing: Beyond Web Accessibility, *Univers Access Hum Comput Interact* **10278** (2017), 322-334.
- [9] S. Lewthwaite, Web accessibility standards and disability: developing critical perspectives on accessibility, *Disabil Rehabil.* **36** (2014).
- [10] S. Schmutz, A. Sonderegger, J. Sauer, Implementing Recommendations From Web Accessibility Guidelines: A Comparative Study of Nondisabled Users and Users With Visual Impairments, *Hum Factors* **59** (2017).
- [11] P. Acosta-Vargas, T. Acosta, S. Luján-Mora, Framework for Accessibility Evaluation of Hospital Websites, International Conference on eDemocracy & eGovernment (2018).

Health Facility Ownership Type and Performance on HIV Indicator Data Reporting in Kenya

Milka B. GESICHO^{a,b,1}, Ankica BABIC^{a,c} and Martin C. WERE^{d,b}

^aDepartment of Information Science and Media Studies, University of Bergen, Norway

^bInstitute of Biomedical Informatics, Moi University, Kenya

^cDepartment of Biomedical Engineering, Linköping University, Sweden

^dVanderbilt University Medical Center, USA

Abstract. In low- and middle-income countries, private and public facilities tend to have highly variable characteristics, which might affect their performance in meeting reporting requirements mandated by ministries of health. There is conflicting evidence on which facility type performs better across various care dimensions, and only few studies exist to evaluate relative performance around nationally-mandated indicator reporting to Ministries of Health. In this study, we evaluated the relationship between facility ownership type and performance on HIV indicator data reporting, using the case of Kenya. We conducted Mann-Whitney U tests using HIV indicator data extracted from years 2011 to 2018 for all the counties in Kenya, from the District Health Information Software 2 (DHIS2). Results from the study reveal that public facilities have statistically significant better performance compared to private facilities, with an exception of year 2017 in reporting of counselling and testing, and prevention of mother-to-child transmission indicator categories.

Keywords. HIV-indicator reporting, health facility, ownership, performance,

1. Introduction

In most LMICs, health facilities are required by the Ministries of Health (MoH) to report on various HIV indicators to aid in monitoring and evaluation of HIV programs, advocacy, policy and decision-making. In general, reported indicators are expected to be timely, accurate and complete. Few studies exist that rigorously evaluate differences in performance, and reviews of performance by facility-type have often led to conflicting conclusions[1]. Ownership of health facilities has the potential to affect the performance of health facilities at meeting these HIV reporting requirements, yet rigorous evaluations on relationship between facility type and HIV indicator reporting are limited. With increasing use of national-level centralized electronic HIV-indicator data aggregation and reporting systems such as the District Health Information Software 2 (DHIS2), data now exists in several LMICs for these evaluations. The aim of this study is to establish the relationship of health facility ownership type with ability to meet HIV indicator

¹ Corresponding Author. Milka Gesicho, E-mail: milcagesicho@gmail.com

reporting requirements, using the case study of Kenya. The reporting requirements assessed in this study are completeness and timeliness in facility reporting.

2. Method

A retrospective observational study was conducted in order to identify the relationship between facility type and performance on HIV indicator reporting in Kenya. Reporting data used was from years 2011-2018. From DHIS2, we extracted the following HIV indicator categories based on Kenya's MoH731² summary form: (i) HIV Counseling and Testing (CT), (ii) Prevention of Mother-to-child Transmission (PMTCT) of HIV, (iii) Care and Treatment (CRT), (iv) Voluntary Medical Male Circumcision (VMMC), (v) Post-Exposure Prophylaxis (PEP), and (vi) Blood Safety (BS). Mann-Whitney U tests were conducted in order to compare the two ownership types.

3. Results

There were no statistical significances in performance in reporting of VMMC and BS indicators, which are peripherally associated with the HIV programs. Statistically significant results in performance, and mean ranks for both private and public varied in the different indicators. Performance in completeness and timeliness also varied by year, with public institutions performing better across multiple indicators with the exception of year 2017 in reporting of HIV CT, PMTC indicators.

4. Discussion

Our study only looks at yearly dimensions, but further analyses could be done by county, facility level, and facility type. This work highlights the key potential of how aggregate reporting data can be used to inform decision-making. Qualitative studies can further help highlight factors promoting or hindering quality indicator reporting

5. Conclusion

In this study, we observed a general trend of public facilities in outperforming private facilities in timeliness and completeness of health facility reporting of nationally-mandated HIV indicators.

References

- [1] S. Basu, J. Andrews, S. Kishore, R. Panjabi, and D. Stuckler, "Comparative performance of private and public healthcare systems in low- and middle-income countries: A systematic review," *PLoS Med* **9** (2012), 19.

² *Comprehensive HIV/AIDS Facility Reporting Form*

Hierarchical Clustering for Image Classification in Dermatology: Towards Mobile Deploying

Miruna D. CIULU^a, Stefan HOLBAN^a, and Diana LUNGEANU^{b,1}

^a*Faculty of Computer Science, Polytechnic University of Timisoara, Romania*

^b*Department of Functional Sciences, University of Medicine and Pharmacy Timisoara, Romania*

Abstract. In the context of increasing interest in computer-assisted diagnosis for skin lesion images and mobile applications to be used in real life settings, we propose a combined desktop-smartphone solution for dermatological image classification. Hierarchical agglomerative and divisive clustering are both implemented as methods of cluster analysis, with the RGB color histogram as descriptor for a global image analysis. The cosine similarity is employed for classifying the query image in one of the available clusters, characterized by their centroids. The solution has been tested with a public database of dermoscopic images, with an overall accuracy of 0.73, 95%CI (0.58;0.85).

Keywords. Decision support, skin cancer, melanoma, color histogram analysis, artificial intelligence

1. Introduction

Skin cancer is an increasingly worrying medical problem and artificial intelligence solutions have proved increasing performance in the recent approaches for cancer diagnosis: clustering in general or for dermatology in particular, support-vector machines, deep learning, or even combinations with successful human-machine partnerships [1,2]. We propose a combined approach of desktop-mobile applications, employing hierarchical clustering for image classification. This is an ongoing work we have tested with a research database of annotated dermoscopic images, PH2 [3].

2. Methods and implementation

Two types of hierarchical clustering methods were implemented: (a) agglomerative bottom-up approach, with three linkage criteria (minimum or single-linkage; maximum or complete-linkage; and average linkage); (b) divisive top-down approach. Two metrics to measure the dissimilarity between the sets of images were implemented: the Euclidean and the Manhattan distances. Figure 1 shows the structure of the application.

¹ Corresponding Author. Diana LUNGEANU, Department of Functional Sciences, University of Medicine and Pharmacy, P-ta E. Murgu 2, 300041 Timisoara, Romania; E-mail: dlungeanu@umft.ro; Tel/Fax +40-256-490288

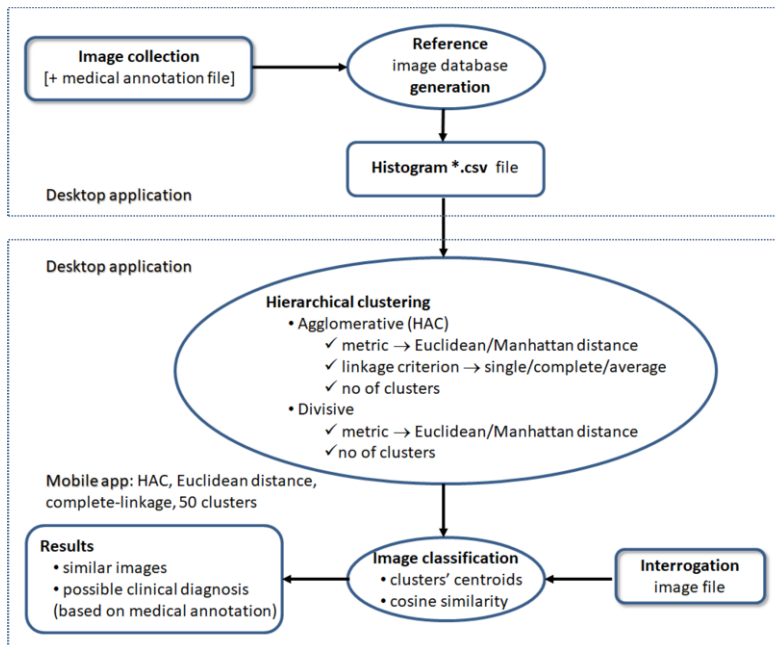


Figure 1. Implementation of the hierarchical clustering: the reference database and histogram file are generated by the desktop application, while the clustering and classification can be performed within the desktop (full options) or mobile (reduced options) application.

3. Results, discussion, and conclusion

The performance assessment was four-folded: (i) hierarchical agglomerative (HAC) vs. divisive clustering; (ii) Euclidean vs. Manhattan distance; (iii) no of clusters; (iv) HAC-related three linkage criteria. The overall results were: sensitivity=0.8, 95%CI (0.61;0.92); specificity=0.6, 95%CI (0.33;0.83); accuracy=0.73, 95%CI (0.58;0.85).

The combined approach we propose merges the advantages of desktop applications' computational power and mobile implementations' friendliness towards a large variety of potential users. The hierarchical clustering presented herewith only uses the global color properties of images, so the classification is generalizable to both dermoscopic and general dermatological images.

References

- [1] S.J. Gilmore, Automated decision support in melanocytic lesion management, *PLoS One* **13** (2018), e0203459.
- [2] A. Hekler, J.S. Utikal, A.H. Enk, A. Hauschild, M. Weichenthal, R.C. Maron, C. Berking, S. Haferkamp, J. Klode, D. Schadendorf, B. Schilling, T. Holland-Letz, B. Izar, C. von Kalle, S. Fröhling, T.J. Brinker, and Collaborators. Superior skin cancer classification by the combination of human and artificial intelligence, *Eur J Cancer* **120** (2019), 114-121.
- [3] T. Mendonca, P.M. Ferreira, J.S. Marques, A.R.S. Rozeira, PH2 - A dermoscopic image database for research and benchmarking, *35th Conf Proc IEEE Eng Med Biol Soc* (2013), 5437-5440 (doi 10.1109/EMBC.2013.6610779).

How to Decide upon Nursing Technologies – A Participation-Based Approach

Jörn KRÜCKEBERG^{a,1}, Ronny KLAUWUNN^b, Inessa FUGE^a,
Reza MAZHARI^c, Regina SCHMEER^c and Nicole HECHTEL^a

^a*Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig –
Institute of Technology and Hannover Medical School*

^b*Institute for Epidemiology, Social Medicine and Health Systems Research, Hannover
Medical School*

^c*Department of Nursing, Nursing Science, Hannover Medical School*

Abstract. This study describes an approach to support decisions on the acquisition of innovative nursing technologies. The approach focusses a participatory design that involves nursing staff from the beginning of the process and aims at achieving positive results regarding identification with the decision of implementation and use of technologies by nursing staff.

Keywords. Nursing technology, participatory management, nursing informatics

1. Introduction

Within the project *center of implementing nursing care innovations* (PPZ Hannover)² innovative technologies are being integrated into everyday nursing work life to support nursing staff and improve patient care on a trauma surgery ward of Hannover Medical School. In order to involve the nursing staff into the research project, a participatory introduction concept was designed and implemented.

2. Methods

Next to an extensive baseline investigation in a mixed methods design [1], the participatory approach uses results from workshops with nursing staff to assess the current work situation on the ward and to derive the need for technical solutions [2]. These results again lead to workshops where innovative nursing technologies are presented and discussed with all nurses of the ward. In the reported workshop a mobilization system for prophylaxis of decubitus ulcer and a robotic companion for dementia care were presented. Instruments for data collection in this setting were

¹ Corresponding Author: E-mail: joern.krueckeberg@plri.de

² Funded by the German Ministry of Education and Research, grant number: 16SV7892K

- guideline-based discussions on usefulness
- a modified questionnaire on the suitability for use [3].

The discussion was documented and analyzed following qualitative content analysis. The results of the questionnaire were calculated for two categories and a total score on defined suitability scores.

3. Results

Both nursing technologies are believed to reduce burdens on patients and nursing staff, while concerns differed between technologies as stated by the nursing staff. Possible improvements for a better adaptation on local conditions were discussed as well. The results of the questionnaire on suitability state a clear tendency for both the products being rated “very suitable” for nursing work life (see table 1). Even though the robotic companion received some criticism in general and in specific aspects of suitability.

Table 1. Summarized results of questionnaire on suitability for use

	General suitability	Specific suitability		
Categories	functionality usability workflow compatibility product quality patient's well-being	power supply alarms (MS) screen (MS) mobility consumables (RC) processing	Suitability for use / total score	
Robotic Companion (RC) n = 15	14 % (3)	19 % (4)	9 % (2)	not suitable
	22 % (4)	31 % (5)	34 % (6)	limited suitable
	64 % (8)	50 % (6)	57 % (7)	very suitable
Mobilization System (MS) n = 14	0 %	12 % (3)	5 % (1)	not suitable
	11 % (2)	33 % (5)	31 % (5)	limited suitable
	89 % (12)	55 % (6)	64 % (8)	very suitable

4. Discussion

The mix of quantitative and qualitative instruments seem to be an appropriate and reliable method to assess suitability of use for nursing technologies. Results of the discussion are matching the results of the questionnaire. The presentation of these results to nursing staff is necessary to comply with the participatory process and might lead to greater identification with the decision of implementation and use of technologies.

References

[1] M. Rutz, R. Schmeer, J. Krückeberg, I. Meyenburg-Altward, M.L. Dierks: PPZ-Hannover - Methodische Überlegungen zur Baseline-Erhebung. In: Boll S, Hein A, et al., eds. Zukunft der Pflege: Tagungsband der Clusterkonferenz 2018. 2018

[2] J. Krückeberg, M. Rutz, H. Hagen, N. Hechtel: Die Perspektive der Basis - Welche Bereiche in der stationären Pflege können von technischen Innovationen profitieren? In: Zukunft der Pflege: Tagungsband der 2. Clusterkonferenz 2019. 2019

[3] L. Müller, C. Backhaus, Entwicklung eines Fragebogens zur Bewertung der Gebrauchstauglichkeit von Medizinprodukten im Beschaffungsprozess, GfA Dortmund (Ed.): Frühjahrskongress Dresden 2019

How to Extract and Explore Big Data for Fraud Detection in the Healthcare Sector: The EOPYY Case Study

Spiros GEORGAKOPOULOS^{a,1}, Parisi GALLOS^a and Vassilis PLAGIANAKOS^{a,b}

^a *Hellenic National Organization for the Provision of Health Services, Greece*

^b *European Healthcare Fraud and Corruption Network, Belgium*

Abstract. Big Data technologies can contribute to medical fraud detection. The aim of this paper is to present the methodological approach of the Hellenic National Organization for the Provision of Health Services (EOPYY) in data analysis to detect financial or medical fraud. To analyze the data for fraud detection, a selection of prescription data from the year 2018 were examined. The Local Correlation Integral algorithm was applied to detect any outliers on the dataset. The results revealed that 7 out of 879 cases could be characterized as outliers. These outliers must be further investigated to determine if they have been associated with fraud. According to the results of this study, this outliers' detection approach can support and help the fraud detection process conducted by the auditing services in Healthcare sector.

Keywords. Fraud Detection, Healthcare, Big Data, Outlier Detection

1. Introduction

Big Data Technologies have been applied in Healthcare sector providing significant advantages on both the quality of the healthcare services and the control of the cost [1]. According to the international literature, the ability to detect fraud in medical claims seems to be a major asset on the expenditure control [2]. Specific Big Data analytic techniques can lead to effective fraud detection in the health domain [3]. The aim of this paper is to present by an example the methodological approach of the Hellenic National Organization for the Provision of Health Services (EOPYY) in data analysis to detect financial or medical fraud in claims.

2. The EOPYY Case Study

EOPYY carries the healthcare provision and claims data of all the Greek citizens. Currently, the databases size is approximately 5 TB with a data increase rate of 140 GB per month, creating a valuable Big Data repository for medical claims. To analyze the data for fraud detection, the 2018 prescription data of the following product categories

¹Corresponding Author, Spiros Georgakopoulos, MSc, PhD, Advisor to the President of the Hellenic National Organization for the Provision of Health Services, Apostolou Pavlou 12, Marousi, Greece; E-mail: spirosgorg@uth.gr.

were extracted (more than 4,000,000 entries) from the above databases: (i) Products and Supplies for Diabetes Management, (ii) Catheters, (iii) Medical Pads, and (iv) Special Nutrition Products and Supplements. For each of the above case, the date of the prescription, the prescriber, the specialty of the prescriber, the patient's social security number, the product name, the category of the product, the product provider, and the location of the provider were selected. After data selection, appropriate anonymization techniques (SHA-1 algorithm for one-way data encryption) were applied to comply with GDPR requirements before the data analysis. The aforementioned raw data were transformed in a manner to describe new features for each product based on the number of the cases. Specifically, the following features were calculated: (i) the number of the prescriptions, (ii) the number of patients who received each product, (iii) the number of prescribers who prescribe each product, (iv) the number of prescription to the number of patients ratio, (v) the number of prescription to the number of prescribers ratio, (vi) the number of the maximum prescriptions per patient, and (vii) the number of the maximum prescriptions per prescriber. A dataset of 879 products with 7 characteristics for each product was constructed. According to the international literature the Local Correlation Integral (LOCI) algorithm [4] is one of the latest algorithms for outliers' detection giving more accurate results on several fields of study [5] compared to other algorithms. Based on this, LOCI was applied to detect any outliers on the dataset, setting the user selected parameter to default values, i.e. $\alpha=0.5$, Nearest Neighbors=20, and Threshold=3, producing some fruitful results. The results revealed that 7 out of 879 products could be characterized as outliers on this seven dimensional space. All of them had different behavior and space distribution, being further away from the rest of the data representing the other products. The 7 outliers may be assumed that have been associated with fraud, since at least two of them have been already confirmed as such by the EOPYY's Auditing Services and the rest are still under investigation.

3. Conclusions

According to the results of this study, the applied outlier detection approach can support and help the fraud detection process conducted by the auditing services in the Healthcare sector. Future work will include the application of multiple algorithms on the healthcare expenditures datasets to achieve more accurate results, while it will also be expanded to the detection of frauds on the prescribers' level.

References

- [1] U. Srinivasan, B. Arunasalam, Leveraging big data analytics to reduce healthcare costs, *IT professional* **15**(6) (2013), 21-28.
- [2] Y. Shi, C. Sun, Q. Li, L. Cui, H. Yu, C. Miao, A fraud resilient medical insurance claim system, *In Thirtieth AAAI Conference on Artificial Intelligence* (2016).
- [3] R. Bauder, T. Khoshgoftaar, Medicare fraud detection using random forest with class imbalanced big data, *In 2018 IEEE International conference on information reuse and integration (IRI)* (2018), 80-87.
- [4] S. Papadimitriou, H. Kitagawa, P.B. Gibbons and C. Faloutsos, LOCI: fast outlier detection using the local correlation integral, *Proceedings 19th International Conference on Data Engineering* (Cat. No.03CH37405), Bangalore, India (2003), 315-326.
- [5] S. Sadik, L. Gruenwald, E. Leal, In pursuit of outliers in multi-dimensional data streams, *In 2016 IEEE International Conference on Big Data (Big Data)* (2016), 512-521.

Identification of EMCI in MR Brainstem Structure Using Fractal Measures and Random Forest Approach

Rohini PALANISAMY¹ and Ramakrishnan SWAMINATHAN

Department of Applied Mechanics, Indian Institute of Technology Madras, India.

Abstract. In this work, an attempt has been made to assess the texture variations of brainstem occurring in the Early Mild Cognitive Impairment (EMCI) condition. Fractal dimension is calculated for the segmented brainstem volume using Higuchi's and Detrending moving average (DMA) method. These measures are validated using Random forest classifier. DMA shows better performance when compared to Higuchi's method in categorising EMCI stage. Thus, the proposed approach using brainstem image's DMA fractal measures with random forest can be used in diagnosis of EMCI.

Keywords. Early Mild Cognitive Impairment, Brainstem, Higuchi's Fractal analysis, Detrended Moving Average, Random forest

1. Introduction

Early Mild Cognitive Impairment (EMCI) describes the clinically asymptomatic stage of Alzheimer's disease (AD). Research on neuropsychiatric symptoms presented by AD patients suggests the early changes in brainstem image texture. Fractals are proven to be effective measures to describe the texture changes in brain tissue.

2. Methodology

Raw T1-weighted age- and gender-matched MR images comprising of 232 EMCI, 200 normal and 200 AD subjects are considered from ADNI database. The images are preprocessed and brainstem volume is segmented using fuzzy based connected component technique. Fractal dimension (FD) is evaluated using Higuchi's method and Detrended Moving Average (DMA) technique. Higuchi's method estimates FD using gray values of pixels and their neighbors. DMA is the advancement of Higuchi's method where, the trend present in the image gray values are removed using root mean square function and the FD values are estimated on the detrended image volume. The extracted fractal measures are further validated using random forest in classifying EMCI stage.

3. Results and Discussion

Representative raw and processed images are shown in Figure 1. The raw images in Figure 1(a) are spatially registered, skull stripped and contrast enhanced as shown in Figure 1(b) and Figure 1(c) respectively. Brainstem volume in Figure 1(f) is segmented from the white matter tissue shown in Figure 1(d). Figure 2(a) shows the linear fit of double log plot calculated using Higuchi's algorithm. The variations in the fluctuation function for different scale obtained using DMA technique is shown in Figure 2(b). The

¹ Corresponding Author, E-mail: rohini2678@gmail.com

slope of double log plot is found to be different for normal, EMCI and AD condition, stating the variations in the fractal nature of brainstem structures. Figure 2(c) shows the box plot indicating mean and standard deviation of fractal values for EMCI, normal and AD subjects. It can be seen that the mean of fractal values calculated using DMA is well separated when compared to Higuchi's method and shows significant differences in distinguishing EMCI condition. As the fractal values are indirect measure of texture heterogeneity, it can be seen that the brainstem images of EMCI patients are comparatively less heterogeneous than normal and AD subjects. This may be due to changes in the neuronal structure during the preclinical stages of the disease. The changes in FD values are further validated using random forest classifier where, an accuracy of 89% is achieved using DMA in classifying the EMCI stage.

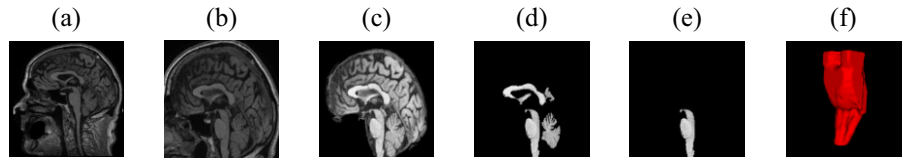


Figure 1. Representative images of (a) raw (b) spatial normalized (c) skull stripped (d) white matter tissue (e) segmented brainstem structure (f) 3D brainstem volume

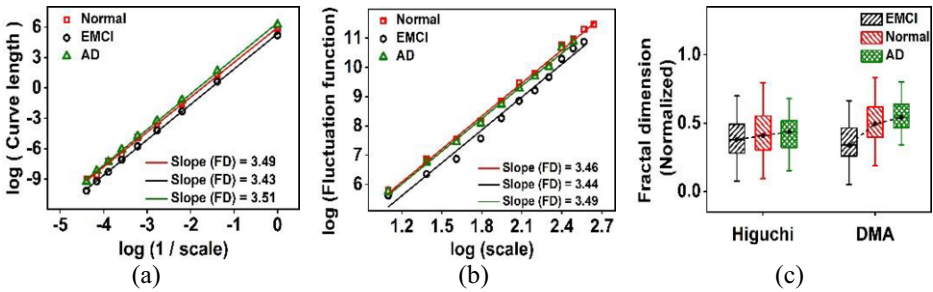


Figure 2. Double logarithmic plot representing FD calculated using (a) Higuchi's and (b) DMA method (c) Box plot representing FD changes among EMCI, normal and AD

4. Conclusion

FD is able to capture the subtle changes in the fractal nature of the brainstem images. The variations in FD obtained using DMA for brainstem volume are found to be significant in differentiating EMCI condition. Brainstem image's fractal measure is found to be distinct in defining the clinical pathology and classifying the EMCI stage, and hence can be used for preclinical diagnosis of AD.

References

[1] J.H. Lee et al., Brainstem morphological changes in Alzheimer's disease. *Neuroreport* **26**(7) (2015), 411.
[2] C. R. Jack et al., Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology* **62**(4) (2004), 591-600.
[3] P. Rohini et al., Characterization of Alzheimer conditions in MR images using volumetric and sagittal brainstem texture features. *Computer Methods and Programs in Biomedicine* **173** (2019), 147-155.

Identifying and Assessing Competencies for Staff Working in Public Health Emergencies

Heini UTUNEN^a, Andrew BLACK^a, Oliver STUCKE^a, Melissa ATTIAS^a, Gaya M GAMHEWAGE^a

^a *World Health Organization*

Abstract. Competency-based learning involves identifying the knowledge, skills and attributes required for carrying out workplace roles. In 2018, the World Health Organization's (WHO) Health Emergencies Programme (WHE) introduced a competency framework with competency-based assessment (CBA) methodology. The CBA focuses on behavioural indicators that participants demonstrate and training faculty observe during training events. This article introduces the CBA of WHE. It is essential to use the framework to design learning programmes and provide the basis to appraise and manage WHE personnel and team performance.

Keywords. Competencies, competency assessment, competency-based learning, learning and development, capacity building, health emergencies

1. Introduction

WHO has set a target of 1 billion more people better protected from health emergencies. To achieve that, a health workforce of excellence is needed, supported by a learning strategy that provides a framework for all learning using a competency-based approach.

Competency-based learning has been derived from Human Resources practices of identifying the knowledge, skills and attributes required for staff to complete the roles assigned to them and linking these roles to the goals of the organization [1]. Employers have made the distinction between core competencies that should be demonstrated by all employees and functional or technical competencies that are required for staff to perform specific roles [2]. In emergency work where staff and leaders face unique pressures, their ability to demonstrate the core competencies becomes particularly important [3]. WHE has followed this model in the roll-out of its competency framework.

2. Method

WHE core competencies are a subset of WHO competencies: Moving forward in a changing environment, applying technical expertise, communication, teamwork, building and promoting partnerships, and leadership. The competencies are demonstrated by 49 observable behaviours. WHE initiated CBA to determine how competencies are met by staff undergoing training and simulation exercises. Participants are assessed by observers and scored based on their demonstration of the behaviours. The scores are aggregated to indicate areas of strength and weakness for each participant and of the whole training cohort. Competency-based discussions are held between

participants and training faculty before and after a training event. These discussions are designed to help participants reflect on the behaviours they feel they readily demonstrate and those which they would like to develop through further learning.

3. Results

The CBA has been piloted in four training activities with approximately 100 participants. It was found that where participants were observed against all 49 behaviours throughout the entire training period, the assessment provided an indication of strengths and areas for development which were largely in line with the participants' self-assessments. Where the number of behaviours observed was reduced, the assessment provided less indication and was less helpful in guiding out-brief conversations with participants. A greater number of facilitators conducting observations improved the assessment consistency. It was also found that it was necessary to conduct a face-to-face training and briefing for the competency observers. Scores for all participants of a training activity are also aggregated into a gross finding to give an indication of areas of strength and weakness for the entire cohort. These can be addressed in future learning programmes.

4. Discussion

The CBA has provided invaluable input in producing meaningful feedback for learners about their demonstration of core competencies. It is also critical for the training owners to mark how the learning objectives are met and the need for further learning. WHE currently has core competencies that are essential for anyone working for WHO in emergencies. Functional competencies are encouraged to be developed by technical teams in charge of functional areas.

5. Conclusions

Providing objective and constructive feedback on behaviours can help an individual identify and reflect on his/her strengths and areas for development, which is a key part of the learning process. Competencies must be understood by the staff and personnel supporting a learning programme; faculty should be trained in their use for designing, delivering and assessing learning and training activities.

References

- [1] Chartered Institute of Personnel Development. 2019. Competency Factsheet <https://www.cipd.co.uk/knowledge/fundamentals/people/performance/competency-factsheet#6369>
- [2] Pfeifer, J 2013. Crisis Leadership the Art of Adapting to Extreme Events. Harvard Kennedy Business School <https://www.hks.harvard.edu/sites/default/files/centers/research-initiatives/crisisleadership/files/Pfeifer%20Crisis%20Leadership--March%2020%202013.pdf>
- [3] Demiroz, F and Kapacu, N. 2012. The role of Leadership in Managing Emergencies and Disasters. European Journal of Economic and Political Studies. <https://pdfs.semanticscholar.org/709b/e5648904a209dcaeece89230ddc9c4b7472aa.pdf>

Identifying Experts Reasoning in Antibiotic Treatment with Preference Learning

Karima SEDKI^{a,1}, Chaymae LAKRAFLI^a, Jean-Baptiste LAMY^a and Rosy TSOPRA^{b,c}

^a*LIMICS (INSERM U1142), Université Paris 13, Sorbonne Paris Cité, 93017 Bobigny, France UPMC Université Paris 6, Sorbonne Universités, Paris*

^b*INSERM U1138 Team22 Information Science to Support Personalized Medicine, Paris*

^c*LTSI Université-Rennes 1 Team Health Big Data, Rennes*

Abstract. The aim of this paper is to propose an approach based on preference learning for building a model that represents the closest possible experts reasoning to provide recommendations. We apply this method to antibiotherapy in primary care. The preference model was learned from the recommendations and from a database describing the domain.

Keywords. Preferences learning, Experts reasoning, Clinical Practice Guidelines

1. Introduction

Clinical practice guidelines (CPGs) are textual documents written by a group of experts. They contain recommendations and their justifications. In the domain of antibiotic prescription, we analyzed these justifications and showed that experts used antibiotic properties to recommend antibiotics [1,2]. The preference model used by experts is based on these properties, but still need to be explicitated, before to be implemented in a clinical decision support system (CDSS). Here, we propose a method for learning preferences [3,4] of CPG experts from both recommendations and a database containing the antibiotics and their properties. We aim at building a model that represents the closest experts reasoning to provide recommendations in the domain of antibiotherapy in primary care.

2. Methods

We used a database containing the antibiotics, their properties, and their rank of recommendations as defined in CPGs. This database was validated by antibiotic experts according to a Delphi process [1]. Let P be the set of properties of antibiotics of the database D . For each property $p_i \in P$, we denote the domaine of p_i by $dom(p_i) = \{1= True, 0=False\}$. Each instance d in D has a recommendation rank. Antibiotics having $R_d=1$ are preferred to those having $R_d=2$ which are also preferred to those having $R_d=3$ which are also pre-

¹Corresponding Author: Karima Sedki, E-mail: karima.sedki@univ-paris13.

ferred to those having $R_d=4$. Antibiotics having R_0 are not preferred since that are not recommended. The aim is to learn preference rules from antibiotics properties, which will allow to associate the same recommendation ranks that those defined in CPGs. Let us give in the following definitions.

Definition 1 (Item) *Each element that is built on the connective logic \wedge and contains at most two properties $p_i \in P$ is called an item. The set of items is denoted by I . We define $Sat(d \in D, x \in X)$ a function that returns True if instance d satisfies item x .*

Definition 2 (Preference rule) *A preference rule r is a partial order in the form of $x_1 \succ x_2 \succ x_3 \succ x_4$ where x_i is an item defined in Definition 1.*

The aim is to define the best items. Hence, we define in the following the support of each item regarding each recommendation rank.

Definition 3 (Support) *The support of an item $x \in X$ regarding instances $d \in D$ having a recommendation rank $R_d = r$ is defined as:*

$$Supp(x, r) = \frac{|\{d \mid Sat(d, x) \text{ and } R_d = r\}|}{|\{d \mid R_d = r\}|}$$

Computing the support of each item regarding each rank allows to determine which item is best for that rank. The interest of an item increases with its support. Our method is summarized as follows:

1. Generate the list of possible items I following Definition 1.
2. Compute the support of each one regarding each recommendation rank R .
3. Determine best items ($BestR_i$) for each R . They are those having the support superior than the fixed minimal support ($thresholdSupp$ is in the $]0, 1]$). Different preference formulas can be generated.

3. Conclusion

The aim of this paper is to propose an approach based on preference learning for building a model that represents the closest possible experts reasoning and strategies to provide recommendations. In the future, we aim at validating our approach in other infectious diseases, and then implementing the preferences rules in AntibioHelp® a CDSS for antibiotic prescription in primary care.

Funding: Agence Nationale de Sécurité du Médicament et des Produits de santé (ANSM). [AAP 2016-RaMiPA Project - Reasoning for a better antibiotic prescription - Dr Rosy TSOPRA]

References

- [1] R. Tsopra, J.B. Lamy, and K. Sedki. Using preference learning for detecting inconsistencies in clinical practice guidelines: Methods and application to antibiotherapy. *AIM*, **89** (2018), 24–33.
- [2] R. Tsopra, K. Sedki, M. Courtine, H. Falcoff, A. De Beco, R. Madar, F. Mechai, J.B. Lamy. Helping GPs to extrapolate guideline recommendations to patients for whom there are no explicit recommendations, through the visualization of drug properties. The example of AntibioHelp in bacterial diseases. *JAMIA*, ocz057, <https://doi.org/10.1093/jamia/ocz057>, 2019.
- [3] J. Frnkranz and E. Hllermeier. *Preference Learning*. Springer-Verlag New York, 1st edition, 2010.
- [4] O. Dekel, C.D. Manning and Y. Singer. Log-linear models for label ranking. In *NIPS*, 497–504, 2003.

Identifying Indicators to Assess and Monitor Data Integration Engines Systems

Priscila MARANHÃO^a, Ana Margarida PEREIRA^a, Ricardo J. CRUZ-CORREIA^{a,b}

^a*Center for Health and Technology and Services Research - CINTESIS, University of Porto, Portugal*

^b*HealthySystems, Portugal*

Abstract. We aimed to identify relevant indicators for end-users in integration engines for healthcare systems. Methods: The study was performed in two steps, including interviews and the identification of additional indicators from the literature. Results: 10 interviews were performed and 90 indicators identified. Discussion: Several of the indicators are difficult to calculate, nevertheless, they have the potential to improve data quality and processes in healthcare institutions and should be further explored in future studies.

Keywords. Integration, indicators, interoperability

1. Introduction

Integration engines are critical in the transmission of clinical and demographic data, maintaining accessibility and integrity within healthcare institutions worldwide [1]. Indicators to assess and monitor data integration engines are essential, contributing both to improve data quality and to facilitate systems management [2]. This study aims to identify relevant performance indicators for end-users regarding the implementation of integration engines.

2. Methods

Data were collected using a 2-step exploratory qualitative approach. The first step was an individual interview with expert professionals from health institutions/health companies that used health information systems daily/almost daily. They were asked to identify indicators within four pre-selected relevant classes of indicators (identified by the authors based on literature review and personal expertise): Information System integration, Data quality, Performance, and Safety and General Data Protection Regulation (GDPR). Two new categories were identified during the interviews: Integration Management and Hospital workflow process. The second step was the identification of additional indicators from the literature [3,4].

3. Results

We performed 10 interviews and identified 63 indicators. Additional indicators (n=27) were identified in the reports. Table 1 presents a sample of the identified indicators by domain. A list of all indicators (n=90) is available at <http://shorturl.at/sIV28>.

Table 1. Indicators concerning integrated software systems per domain; only a sample is presented.

Information System integration	• Absolute n of messages delivered / n sent	• Calculate n of alerts by period of time
Data quality	• Data availability	• Total messaging problems
Performance	• Delivery/receiving time	• N of SLA metrics for integration
Safety and GDPR	• Access history checking	• Key expiration time
Integration Management	• N of integration problems	• N of hours to set up/manage
Hospital work process indicators	• Number of coding errors	• Process reengineering

N – number; SLA – Service Level Agreement; GDPR – General Data Protection Regulation

4. Discussion

We believe that this study describes a relevant list of indicators that should be assessed and promoted within an integrated health system. Although several of the indicators are difficult to compute, they might contribute to improve data quality and processes in healthcare institutions, by promoting the identification and facilitating the correction of system inefficiencies or malfunctions, and should be further explored in future studies.

5. Conclusion

This study identified 90 indicators that can be analyzed by actors involved in the healthcare system and have the potential to contribute to data quality improvement.

References

[1] G.D. Armitage et al. Health systems integration: state of the evidence, *International Journal of integrated Care* **9** (2009), 1-11.

[2] J.Braithwaite et al., Health system frameworks and performance indicators in eight countries: A comparative international analysis, *SAGE open Medicine* **5** (2017), 1-10.

[3] Mordor Intelligence. Global Healthcare IT Integration Market. Available: <https://www.mordorintelligence.com/>. 2017

[4] HL7 Interface Engine Ratings - CoreHealth. Available: <https://www.corehealthtechnologies.com/2018-hl7-interface-engine-comparison-ratings/>. 2018

Implementation of Wireless Biosensor for Continuous EEG Monitoring in Neurological Intensive Care

Tien-Yu HSU^a, Terry B.J. KUO^b, Ming-Chin LIN^{c,1}

^aGraduate Institute of Biomedical Informatics, College of Medicine Science and Technology, Taipei Medical University, Taipei, Taiwan

^bInstitute of Brain Science, National Yang-Ming University, Taipei, Taiwan

^cGraduate Institute of Biomedical Informatics, College of Medicine Science and Technology, Taipei Medical University, Taipei, Taiwan

Department of Neurosurgery, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan

Taipei Neuroscience Institute, Taipei Medical University, Taipei, Taiwan

Abstract. Seizure is a common complication in a neurological intensive care unit (NICU) and it requires continuous electroencephalograms (EEG) monitoring. Implementation of EEG for each bed in a NICU is very expensive and require labor work for interpretation of EEG. To provide an affordable device of EEG in NICU, we developed a low-cost wireless biosensor, which utilized the current standard of the internet of things technology (IoT). In this study, we implement a wireless biosensor for continuous EEG monitoring in NICU and discuss its feasibility. To provide a low-cost EEG device, we embraced Bluetooth and mobile phone technology, which is convenient for implementation. We build a two-channel EEG biosensor, which utilizes Bluetooth to transmit the signal to mobile phones. Then, mobile phones use Wi-Fi technology to send data to the server. Additionally, we also developed a registry to organize the patient's EEG data. In six months research period, we have 65.8% of patients collected successfully. Using 2 channel-biosensor in NCU is feasible. It also develops a neuromedical database by collecting and monitoring physiological signals to develop future neuromedical research.

Keywords. Continuous EEG, two-channel biosensor, Neurointensive care unit, wearable device

1. Introduction

Seizure is a common complication in a neurological intensive care unit (NICU), which the incident rate ranges from 3%-45% [1] and it requires continuous electroencephalograms (EEG)[2] monitoring. Implementation of EEG for each bed in a NICU is very expensive and require labor work for interpretation of EEG. To provide an affordable device of EEG in NICU, we developed a low-cost wireless biosensor[3-4], which utilized the current standard of the internet of things technology (IoT). In this study, we implement a wireless biosensor for continuous EEG monitoring in NICU and discuss its feasibility.

¹Corresponding author: E-mail: arbiterski@gmail.com

2. Method

To provide a low-cost EEG device, we embraced Bluetooth and mobile phone technology, which is convenient for implementation. We build a two-channel EEG biosensor, which utilizes Bluetooth to transmit the signal to mobile phones. Then, mobile phones use Wi-Fi technology to send data to the server. Additionally, we also developed a registry to organize the patient's EEG data.

3. Result

From April 1, 2019 to September 30, 2019, there were 369 patients and 42 patients excluded. After deducting repeat patients, the number of cases was 241. The age ranged from 16 to 96 years. The main diagnoses included intracranial hemorrhage (43.9%), infarction(24.1%), subdural hemorrhage (15.1%), subarachnoid hemorrhage (14.0%), brain tumor (9.2%), aneurysm (8.6%), seizure(6.2%) and epidural hemorrhage (4.3%). The male to female ratio was 58.3%:41.7%. We reviewed the EEG data in the database and then compared them with nursing records and found that they were consistent with the seizure attack time.

4. Discussion:

In six months research period, we collected 65.8% of NICU patients successfully. Those patients which cannot be collected were due to 1.Network connection (16.2%) 2.Patient skin irritation (11.9%), 3.Irritable patients (8.1%), 4.Wireless signal(6.5%), 5.Operation wound (3.2%), 6.Out of battery (1.2%). To improve the above issues, we need to design a better server program that can stably pass data. For the skin irritation, we try to use some field for skin protection. Irritable patients will be excluded until the condition is stable. To enhance the wireless signal, the Bluetooth connective firmware needs upgrades and the mobile phone can be closer to the biosensor. The operation wound some times would overlapping on where we stick the sensor lead. To prevent influenced the treatment and infection, we exclude these cases. The battery wears in NICU stronger another place because vapor inhalation treatment is often in NICU. During the study period, 7 of the 34 battery circuit boards became rusty. Therefore, the development of a waterproof function is an essential topic in the future.

5. Conclusion

We can use wireless biosensor to collect continuous EEG and there were 65.8% of NICU patients collected successfully. It also develops a database by collecting and monitoring physiological signals to develop future neuromedical research.

References

- [1] S. Yerram, N. Katyal, K. Premkumar, P. Nattanmai, and C.R. Newey, Seizure prophylaxis in the neuroscience intensive care unit, *J Intensive Care* **6** (2018), 17.
- [2] Y. Kubota, H. Nakamoto, S. Egawa, and T. Kawamata, Continuous EEG monitoring in ICU, *J Intensive Care* **6** (2018), 39.
- [3] Y.-C. Su, Exploration of EEG Learning Curves during the Music Playing Process from Rusty to Skilled Performance, (2019).
- [4] M.A. Serhani, M.E. Menshawy, A. Benharref, S. Harous, and A.N. Navaz, New algorithms for processing time-series big EEG data within mobile health monitoring systems, *Comput Methods Programs Biomed* **149** (2017), 79-94.

Implications of Clinicians' Attitudes Towards Clinical Decision Support Systems

Anshu ANKOLEKAR^{a,1}, Britt VAN DER HEIJDEN^a, Andre DEKKER^a, Cheryl ROUMEN^a, Dirk DE RUYSSCHER^a, Bart REYMEN^a, Ruud HOUBEN^a, Peter FICK^a, Sander PUTS^a, Joeri VEUGEN^a, Adriana BERLANGA^b, Cary OBERIJE^c, Rianne FIJTEN^a

^a Department of Radiation Oncology (MAASTRO Clinic), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands

^b Maastricht University, Maastricht, The Netherlands

^c The D-Lab, GROW - School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands

Keywords. Shared decision making, decision aids, clinical decision support, lung cancer

1. Introduction

Clinical decision support systems (CDSSs) can provide personalized predictions for cancer patients, such as overall survival and risk of side-effects associated with different treatments. Combined with shared decision-making (SDM), CDSSs can guide treatment decisions by supporting data-driven decision-making in line with patient values. Prior findings have established the value of SDM and CDSSs in improving care, yet implementation remains a significant hurdle due to barriers such as lack of accessibility, lack of clarity about the tool's purpose, time pressure, and poor design.

The purpose of this study was to explore the potential for implementing a CDSS in combination with SDM in the lung cancer trajectory in the Netherlands from the perspective of lung cancer specialists.

2. Methods

We used a mixed methods design consisting of qualitative interviews with 9 Dutch lung cancer specialists (7 pulmonologists and 2 radiation oncologists) and a retrospective cohort study on MTB treatment decisions of 257 non-small cell lung cancer (NSCLC) stages I-III patients in our clinic. Interview themes included the treatment decision-making process, clinician views on CDSS and their requirements.

¹ Corresponding Author, Department of Radiation Oncology (MAASTRO Clinic), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands; E-mail: anshu.ankolekar@maastro.nl.

In addition, we determined the percentage of lung cancer patients that deviate from MTB recommendations and their reasons for deviation. We also compared the treatment outcomes and quality of life of deviating patients with those of non-deviating patients with chi-square and t-tests respectively.

3. Results

The majority (78%) of the clinicians were receptive towards CDSSs overall but cited barriers such as lack of trust in current prediction models, lack of validation studies, and time pressure. They felt a CDSS would be most valuable during the MTB discussion between clinicians rather than in consultations with patients. Current MTB recommendations are based on clinician experience, clinical guidelines and patients' WHO performance status, and our cohort study revealed that in 89% of cases these recommendations were followed by patients. Indeed, although the specialists were supportive of SDM initiatives, the majority felt that it is difficult to practice in lung cancer as there is often one 'best' treatment option.

Of the patients who deviated from MTB recommendations, 82.1% did so in order to choose a less intensive treatment. We found no significant differences in overall survival and toxicity between the two patient groups and a minor increase in quality of life in the deviating group.

4. Discussion

Our findings have implications for developing and implementing CDSS and SDM in lung cancer treatment. First, MTBs are generally characterized by lack of time, failure to reach decisions, and failure to take patient preferences into account. Second, clinicians' survival predictions are positively biased and radiation oncologists are limited in their ability to predict lung cancer treatment side effects. This contrasts strongly with our finding that clinician experience was the primary factor influencing MTB decisions, and underlines the need for data-driven CDSS.

Furthermore, when clinicians are skeptical of CDSSs due to lack of validation studies, steps can be taken to involve them in a user-centered design process so that their inputs, requirements, and feedback guide CDSS development. Finally, integrating the CDSS into the existing workflow is crucial to minimize the additional time required for clinicians to fill in patient data.

5. Conclusion

There is scope for CDSSs to support decision-making in lung cancer treatment, however care must be taken to incorporate these CDSSs into the clinical workflow.

Incorporating Tacit Knowledge of Experts in the Assessment of Shelters Under Disaster

Masaharu NAKAYAMA, MD, PhD ^{a,1} and Tadashi ISHII, MD, PhD ^b

^aMedical Informatics, Tohoku University Graduate School of Medicine, Miyagi, Japan.

^bDepartment of Education and Support for Community Medicine, Tohoku University Hospital, Miyagi, Japan

Abstract. During disasters, prompt assessment of evacuation centres and appropriate measures taken based on the assessment are important for protecting local residents' lives and health. To clarify experts' tacit knowledge for improving the management of shelters under disaster, we conducted a questionnaire survey of specialists in disaster management to examine the assessment items to which they accorded top priority. Our results suggest that specialists give highest priority to the number of evacuees, followed by infrastructure and need for medical intervention. Identifying their common views on priorities can be the first step towards taking standard appropriate measures in evacuation centres.

Keywords. disaster management, shelter, Great East Japan Earthquake

1. Introduction

The Great East Japan Earthquake (GEJE) and tsunami on March 11, 2011, caused an unprecedented disaster with 15,897 deaths and 2,534 missing persons as of December 10, 2018 [1]. During the disaster, many people (around 400,000) were evacuated to temporary shelters, such as schools, gymnastic halls, and public halls in Northeast Japan. Several organized teams cooperating with disaster specialists, various medical institutions, and local government officials managed to improve the conditions of the shelters [2]. It is important to understand how disaster experts make a decision based on the status of a shelter, which would be useful for all potential managers in future disasters. Our study aimed to clarify the algorithm through a questionnaire survey.

2. Methods

We conducted a questionnaire survey on 11 disaster management experts who had experienced the GEJE in 2011 or the Kumamoto earthquake in 2016. The research protocol was approved by the Tohoku University Graduate School of Medicine Research

¹ Corresponding author: 1-1 Seiryomachi, Aoba-Ku, Sendai, Miyagi 980-8574, Japan;
E-mail: nakayama@cardio.med.tohoku.ac.jp

Ethics Committee. The questionnaire comprised two questions, to examine 1) which of the 30 assessment items is given top priority and 2) whether a facility accommodating 100 people (a larger shelter) with medium urgency or a facility accommodating 10 people (a small shelter) with greater urgency should be given priority, for each assessment item. More respondents favoured the former. Opinions were divided on the importance of the number of diseased/injured people.

3. Results

The number of evacuees received the highest priority, followed by drinking water supply, information about diseased or injured people, basic information, number of diseased or injured people, electricity, sanitation, and information about people who require assistance. Next, to examine whether the priority changes depending on the number of evacuees and severity, we asked respondents to consider which they would give priority to between a shelter accommodating 100 evacuees (a larger shelter) with medium urgency and a shelter accommodating 10 people (a small shelter) with greater urgency for each assessment item. Most of them felt that the larger shelter required priority even though the smaller shelter was in a more serious situation. Among the 30 items, electricity, waterworks, drinking water supply, and food supply were evaluated as relatively important.

4. Discussion

Our results suggest that specialists give highest priority to the number of evacuees, followed by infrastructure and need for medical intervention. Clarifying specialists' common views on priorities is the first step in identifying standard appropriate measures in evacuation centres. The present study has several strengths. First, our results suggest that specialists give highest priority to infrastructure and need for medical intervention. The study presents the tacit knowledge of specialists who are considerably experienced in disaster management. Second, this study also revealed a slightly different priority based on urgency. This implies that information to evaluate the urgency of shelters is crucial at the headquarters when an area is struck by disaster, resulting in shelters needing support using limited resources of manpower, medicines, food, and water. As advance preparation to establish an initial response system and decision-making is critical for disaster response management, this study's findings may be useful in doing so.

References

- [1] National Police Agency, Report on the Damage from the Great East Japan Earthquake, 2018, <https://www.npa.go.jp/news/other/earthquake2011/pdf/higaijokyo.pdf> [Accessed: 24 Jan 2019]
- [2] Ishii, T. The report of the Great East Japan Earthquake relief activities in the Ishinomaki medical district from the perspective of a Miyagi Prefecture disaster medical care coordinator. *J Disast Med.* **17**(2012), 92-98.

Acknowledgements

This study was supported by a Grant-in-aid from the Japanese Ministry of Health, Labour and Welfare, Tokyo, Japan (15K08799). We are extremely thankful to Misaki Arakawa, Mikiko Sato, Kumiko Tamura, and Yukie Kobayashi for the technical support they offered.

Integrating Socially Assistive Robots into Japanese Nursing Care

Markus KOLSTAD ^{a,1}, Natsu YAMAGUCHI ^b, Ankica BABIC ^{a,c} and
Yoko NISHIHARA ^d

^a*Department of Information Science and Media Studies, University of Bergen, Norway*

^b*College of Law, Ritsumeikan University, Japan*

^c*Department of Biomedical Engineering, Linköping University, Sweden*

^d*College of Information Science and Engineering, Ritsumeikan University, Japan*

Abstract. This paper presents experiences of integrating assistive robots in Japanese nursing care through semi-structured interviews and site observations at three nursing homes in Japan during the spring of 2019. The study looked at experiences with the robots *Paro*, *Pepper*, and *Qoobo*. The goal was to investigate and evaluate the current state of using robots in the nursing care context, firsthand experiences with intended and real use, as well as response from the elderly and nursing staff. The qualitative analysis results pointed out user satisfaction, adjusted purpose, therapeutic and entertaining effects. Potentials of robots to assist in elderly care is advantageous. Limitations currently relate to the lack of ways to fully utilized and integrate robots.

Keywords. Assistive Robots, Nursing Care, Human-Robot Interaction (HRI), Therapy, Communication, Impact on Care.

1. Introduction

According to the United Nations (UN) demographic statistics, Japan's population has decreased since 2009. Japan's population today (2019) is about 126,800,000, but the forecast is that the number will fall to 108,800,000 by 2050 [1]. Moreover, Japan currently holds the world's largest population of elderly per capita with about 27% of Japan's population being 65 years or older. In order to cope with the changing demographic, possible solutions are to decrease the need for more human workers through innovation and to keep the elderly as independent as possible for longer.

The Japanese government has presented six priority areas where robot technology is to be introduced in nursing care, one of them being *monitoring and communication systems* [2]. We visited three different Japanese nursing facilities in order to see how robots are being used, what impact they made on the nursing care, and what positive and negative experiences the elderly and staff could share. The interviews focused on communication robots, being one of the newest items introduced as the government's priority areas whose development has been supported since 2017.

¹ Kolstad Markus, Department of Information Science and Media Studies, University of Bergen, Norway;
E-mail: markuskolstad@gmail.com.

2. Methods

The three nursing facilities were chosen as the study sites from a public list of nursing facilities using the robot *Paro* [3]. The interviews were not limited to *Paro* exclusively, but rather aimed at acquiring information on all interactive robots used at each facility (*Paro* was used at all 3 facilities, *Pepper* at two, and *Qoobo* at one). The interviews were primarily held with the facilities' managers and nursing staff in addition to observations of patients interacting with robots. The interviews were analyzed using open coding.

3. Results

The interviews suggested that these kinds of robots all work with the mind and the mental wellbeing of the patients. Patients suffering with dementia often believe *Paro* to be a dog or even a human baby, treating and interacting with it accordingly. Patients often cared and worried for *Paro*, thus indicating an emotional relationship. Interacting with *Paro* evoked memories about their childhood dog or child nursing days. *Paro* was found effective on both male and female patients if they liked animals, but patients that are more cognitively capable might find *Paro* less interesting. The use of robots has also depended on the nursing staff attitude. *Pepper* is used for recreation, entertaining patients through games and karaoke. Still, some patients found it loud and annoying. *Qoobo* was mostly used by one patient not interested in people but happy with the robot.

Discussion

Robots can be superior to the use of dolls or animals as therapeutic aids should be safe and have demonstratable therapeutic effects. Robots like *Paro* can work as an alternative to pet/animal therapy without the risk of allergy, bites, scratches, unpredictability, infections or even stress for the animal itself. However, some people dislike certain animals which makes them skeptical to animal-type robots as well. *Pepper* is mostly used to entertain, in addition to Taiso, the daily Japanese stretching exercise. However, the need for human workers is not eliminated. Ultimately, these robots are used to assist and relieve the humans from work overload. Patients, that usually did not smile to the human staff, were seen smiling and talking to the robots. Additional positive effects were increased interest of groups and often grandchildren to visit the nursing homes.

Conclusion

Positive effects were patient satisfaction, joy, and mood changes. On the downside, robots are not fully independent and demand staff attention which limits their use.

References

- [1] <https://population.un.org/wpp/Graphs/DemographicProfiles/>
- [2] https://www.meti.go.jp/english/press/2017/1012_002.html
- [3] <https://www.daiwahouse.co.jp/robot/paro/map.html>

Introducing Clinical Informatics Course in Medical School Curricula: Lessons Learned from Medical Faculty University of Belgrade

Jelena MILIN-LAZOVIC^a, Andja CIRKOVIC^a, Marko SAVIC^a, Ognjen MILICEVIC^a, Ljubica CAREVIC^c, Nikola ILIC^c, Dejana STANISAVLJEVIC^a and Natasa MILIC^{a,b}

^a*Institute for Medical Statistics and Informatics, Faculty of Medicine University of Belgrade*

^b*Division of Nephrology & Hypertension, Mayo Clinic, Rochester, MN, USA*

^c*Center for Information and Communication Technologies, Faculty of Medicine, University of Belgrade, Belgrade, Serbia,*

Abstract. The healthcare environment in Serbia has changed dramatically over a last two decades, pointing out the necessity of clinical informatics (CI) education for future MDs. Total of 77 students were enrolled and 72 (93.5%) have successfully finished this course during 4 academic years. Mean total score for all students was 83.4 ± 9.0 points, without difference between genders. We presented blended learning module as an effective way of gaining competences in CI and recommend this course to be required for future MDs.

Keywords. clinical informatics, education, blended learning

1. Introduction

The healthcare environment in Serbia has changed dramatically over the last two decades, pointing out the necessity of clinical informatics (CI) education for future MDs [1]. Therefore, our main goal for development and implementation of clinical informatics course was to enable students to better understand the uncertainty of medical data, information and knowledge via blended learning mode as effective strategy in delivering education [2,3].

2. Method

Clinical informatics was first introduced in 2015/16 as an elective course for 6th year medical students of Medical Faculty University of Belgrade. This elective course was among 20 proposed, and a minimal number of students had to preselect it in order to be enrolled. Curriculum was developed to enable students, future MDs, to understand the uncertainty of medical data, information and knowledge. Course

materials included four topics and two projects (Figure 1). The blended teaching approach combined on-site lectures and interactive multimedia didactic materials which were accessed through the Moodle Learning Management System (LMS). Course competences were evaluated through projects, quizzes and knowledge test with maximum total score of 100.

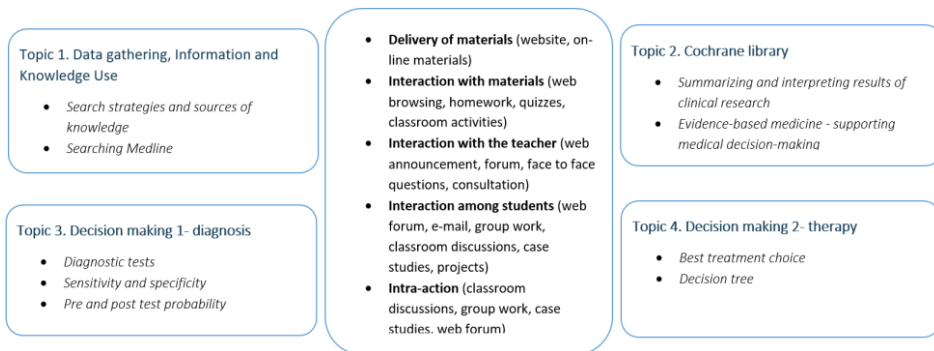


Figure 1. Structure of Clinical informatics Course

3. Results

During the past four academic years, 77 medical students chose Clinical informatics as an elective subject. Seventy-two (93.5%) successfully finished the course. Competences assessment, presented by mean total course score of 83.4 ± 9.0 (for all enrolled students), suggested that students adopted knowledge and skills necessary for medical decision making in conditions of uncertainty. There were no gender differences, as well as no significant changes in scores over a 4-year period.

4. Conclusion

Clinical informatics course is an effective way for medical students to gain competences needed to better understand the medical data uncertainty. In the era of an overwhelming presence of new technologies supporting both learning and health care delivery, clinical informatics should become an obligatory course in medical school curricula.

References

- [1] Milic NM, Ilic N, Stanisavljevic DM, Cirkovic AM, Milin JS, Bukumiric ZM, Milic NV, Savic MD, Ristic SM, Trajkovic GZ. Bridging the gap between informatics and medicine upon medical school entry: Implementing a course on the Applicative Use of ICT. *PLoS One*. 2018 Apr 23;13(4):e0194194
- [2] Sánchez-Mendiola M, Martínez-Franco AI, Rosales-Vega A, Villamar-Chulin J, Gatica-Lara F, García-Durán R, et al. Development and implementation of a biomedical informatics course for medical students: challenges of a large-scale blended-learning program. *J Am Med Inform Assoc*. 2013;20(2):381–7

Knowledge Driven Phenotyping

Honghan WU ^{a,1}, Minhong WANG ^a, Qianyi ZENG ^a, Wenjun CHEN ^a,
Thomas NIND ^a, Emily JEFFERSON ^a, Marion BENNIE ^a, Corri BLACK ^a,
Jeff Z. PAN ^a, Cathie SUDLOW ^a and Dave ROBERTSON ^a

^a*Working Group of Graph-Based Data Federation for Healthcare Data Science (Sprint Exemplar Project funded by Health Data Research, United Kingdom)*

Abstract. Extracting patient phenotypes from routinely collected health data (such as Electronic Health Records) requires translating clinically-sound phenotype definitions into queries/computations executable on the underlying data sources by clinical researchers. This requires significant knowledge and skills to deal with heterogeneous and often *imperfect* data. Translations are time-consuming, error-prone and, most importantly, hard to share and reproduce across different settings. This paper proposes a knowledge driven framework that (1) decouples the specification of phenotype semantics from underlying data sources; (2) can automatically populate and conduct phenotype computations on heterogeneous data spaces. We report preliminary results of deploying this framework on five Scottish health datasets.

Keywords. health data, phenotype computation, data integration, ontology

1. Introduction

Health data in the UK is stored in different local communities, meaning they are maintained locally and stored in inconsistent formats and languages. A key technical challenge haunting almost all clinical studies is to extract or compute accurate patients' phenotypes (traits of symptoms, diseases, medications or biochemistry test results) from such a fragmented data space. Specifying the computations of a phenotype requires (each time) significant human effort to understand database details, good data science skills to do the querying and data manipulating, and caution & patience to deal with data incompleteness/inconsistencies. Such phenotype specifications can hardly be reused as the consequence of the heterogeneous data models across jurisdictions. This significantly impedes the reusability and reproducibility of clinical researches.

2. Method

The main aim of this study is to realise a clinical data science framework [1] that makes the underlying data sources *transparent* to phenotype computations. The key is to decouple the formalisation of phenotype semantics and the technical details of underlying data sources. We propose an architecture to implement such a decoupling, which is composed of the following two aspects.

¹Corresponding Author: 9 Little France Road, Edinburgh EH16 4UX, UK; E-mail: honghan.wu@ed.ac.uk.

Phenotype Formalisation Framework This has three components: (1) A database independent phenotype formalisation using Semantic Web knowledge representation technologies to define phenotype semantics; (2) A core phenotype ontology serving as the base vocabulary linking to standard clinical terminologies available at BioPortal (e.g., SNOMED CT, ICD10); (3) A query formatter that generates ontology queries from an user interface. The formatter can automatically translate phenotype definitions between standard terminologies (e.g., READ to SNOMED CT).

Ontology Based & Rule Driven Data Access To automatically compute the above formalised phenotypes on actual data, we adopt ontology based data access (OBDA) techniques [2] but with a novel extension to support rules. Such an extension is necessary because the semantics of most phenotypes are not fixed. They either change with research focuses or different researchers might have different opinions about certain rules related to a phenotype. For this reason, in our framework, we minimise the ontology. Instead, a **rule engine** is implemented to allow flexible definitions of phenotypes. The engine can automatically convert user specified rules into new data mappings, which will be translated on the fly to do phenotype computations on databases.

3. Deployment and Evaluation

This study was supported by Health Data Research UK (<https://www.hdruk.ac.uk/projects/graph-based-data-federation-for-healthcare-data-science/>) and the Medical Research Council [grant number MC_PC_18029] as an exemplar to create a federation of distributed health data in Scotland. The above described framework has been deployed on 5 synthetic data sets generated using BadMedicine [3], which represents data/schema characteristics learnt from real data. Due to space limitations, we put the full benchmark and evaluation details on a Github page: <https://github.com/Honghan/KGPhenotyping/tree/master/evaluation>.

4. Conclusion

To overcome *obscure* phenotype computation, which makes experiments difficult to understand and reproduce, we developed a new framework to allow clinical researchers to formalise phenotype semantics independently to the data and, more importantly, in a computer understandable way so that its computation can be automated on the underlying data sources. We implemented a knowledge-driven (based on ontologies and rule languages) approach to define an interlingua in which practitioners can represent the phenotype semantics they want to use and automatically translates this to computations as database queries on participating data sources.

References

- [1] Knowledge Driven Phenotyping - an open source Github repository for computing phenotypes across heterogeneous datasets: <https://github.com/Honghan/KGPhenotyping>.
- [2] Xiao, G., Calvanese, D., Kontchakov, R., Lembo, D., Poggi, A., Rosati, R. and Zakharyashev, M., 2018, July. Ontology-based data access: A survey. IJCAI.
- [3] BadMedicine - a synthetic Scottish health data generator: <https://github.com/HicServices/BadMedicine>

Limits and Variability in Drug Databases: Lessons Learnt from Drug Comparisons

Jean-Baptiste LAMY^{a,1}, H  l  ne BERTHELOT^a, Madeleine FAVRE^{b,c}, Rosy
TSOPRA^{d,e}

^aLIMICS, Université Paris 13, Inserm, Sorbonne Université, 93017 Bobigny, France

^bDept. of primary care, Université Paris Descartes, Paris, France

^cSociété de Formation Thérapeutique du Généraliste (SFTG), Paris, France

^dINSERM U1138 Team22 Information Science to Support Personalized Medicine, Paris

^eLTSI Université-Rennes 1 Team Health Big Data, Rennes

Abstract. Several drug databases exist, but sometimes differ in their content. Here, we propose a taxonomy of the variability we observed, and we present a tool for investigating the variability through the visual comparison of the properties of a given drug as represented in several sources or databases.

Keywords. Drug databases, Computer Graphics, Nonverbal Communication

1. Introduction

It is known that drug databases suffer from many discrepancies, both intra- and inter-databases. In a recent project named VIIIP (Integrated Visualization of Information on Pharmaceutical Innovation) [2], we aimed at comparing new drugs to the older ones with the same indication, on the basis of their properties. We planned to extract these properties from drug databases. However, we encountered many problems related to the quality and the structure of these databases, which impaired the project.

2. Methods

We compared the 4 drug database available in France, a Natural Language Processing (NLP) tool [3], and a manual coding by an expert pharmacist (HB), for 10 drug. We used ICD10 and MedDRA terminologies. We analyzed the differences between sources and built a taxonomy of variability. We also used rainbow boxes, a set visualization technique [1] that was initially designed for comparing similar drugs [2]. We adapted it for comparing the properties of the same drug, described by 6 sources.

3. Results and discussion

We identified 3 main causes of variability in drug knowledge:

1) Differences in the writing of SPCs: when two drugs share a common property, it might be described differently in their SPCs. It includes: 1a) Variability in the

presentation of the SPC, e.g. plain text and/or table; 1b) Variability in the file format, e.g. PDF vs HTML; 1c) Variability in the wording; 1d) Variability in the level of granularity (e.g. “liver injury” vs a list of 10 conditions that are the 10 most common liver injuries).

2) Differences in the drug databases: each database has its own structure and editorial practice. It includes: 2a) Missing information (e.g. seriousness of adverse effects); 2b) Structural limitation, when the desired information is present but not structured or coded (e.g. information available in free text only); 2c) Poor information quality; 2d) Heterogeneous coding, when the same information is coded differently for various drugs.

3) Each drug database employs several experts, having their own coding practice.

Figure 1 shows the variability found at level #2. Sources are shown in columns and ordered by similarity. Rectangular boxes indicate the adverse effects. Boxes that span over several columns represent adverse effects present in several sources. Box colors indicate the frequency of the adverse effects, from rare (yellow) to frequent (orange).

In conclusion, it appears that drug databases are designed for providing information on a single drug, but not for comparing two or more drugs.

Funding: This work was partly supported by the French drug agency (ANSM) through the VIIIP project (AAP-2012-013).

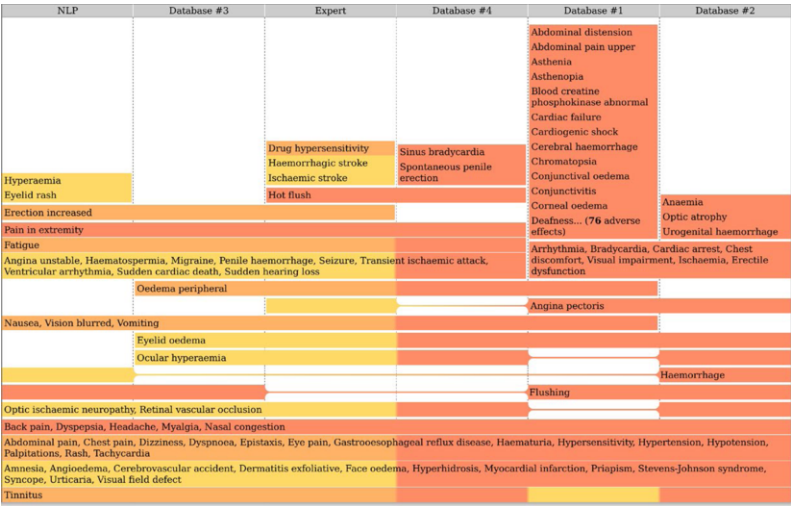


Figure 1. Rainbow boxes comparing the adverse effects of Cialis® from 6 sources.

References

[1] Lamy, J.B., Berthelot, H., Capron, C., Favre, M.: Rainbow boxes: a new technique for overlapping set visualization and two applications in the biomedical domain. *Journal of Visual Language and Computing* **43**, 71–82 (2017)

[2] Lamy, J.B., Berthelot, H., Favre, M., Ugon, A., Duclos, C., Venot, A.: Using visual analytics for presenting comparative information on new drugs. *J Biomed Inform* **71**, 58–69 (2017)

[3] Lamy, J.B., Ugon, A., Berthelot, H.: Automatic extraction of drug adverse effects from product characteristics (SPCs): A text versus table comparison. *Studies in Health Technology and Informatics* **228**, 339–343 (2016)

Linking Genome and Exposome: Computational Analysis of Human Variation in Chemical-Target Interactions

Liana BRUGGEMANN^a, Christopher HAWTHORNE^b, Ram SAMUDRALA^a and Guillermo H. LOPEZ-CAMPOS^{b,1}

^a*Department of Biomedical Informatics, State University of New York (SUNY), Buffalo, NY, USA*

^b*Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, UK*

Abstract. The growing amount of available public data repositories containing a plethora of rich chemical and biomedical information is enabling new in silico research avenues. In this project we aim to link human genome variations and the exposome applying in silico biomedical informatics approaches to analyse the potential effects of those variants in the interactions with different chemicals.

Keywords. Translational bioinformatics, Protein modelling, exposome, single nucleotide variant, toxicogenomics.

1. Introduction

Nowadays, human societies are living in an environment where individuals are exposed to an increasing number of xenobiotic chemicals. The acknowledgment of the importance of these and other exposures led to the development of the concept of the exposome defined as the whole set of exposures of an individual (including, chemical, biological, physical or and social exposures) [1]. The exposome could affect the human phenotype along with the genome and therefore its understanding could be important in the development of precision medicine.

Two decades ago, toxicogenomics was developed combining toxicology with genomics and post-genomics approaches to study the effects of the exposures to chemicals and chemical toxicants. Biomedical informatics has played a key role in this discipline, providing tools and methods to analyse, store and integrate the results from these “omics” techniques. More recently, the development of modelling and simulation tools and methods designed to predict the potential interactions of newly designed chemicals with biological targets (mostly hormone and other cellular receptors) has been another important area of development. Toxicogenomics can be considered as the toxicological relative of pharmacogenomics. An important area of research in the latter is the analysis of the effects of the genetic variants in the response to different drugs, and this has led to different applications and biomedical informatics resources. The analysis

¹ Corresponding Author, Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, 97 Lisburn road, Belfast, BT9 7BL, United Kingdom; E-mail: g.lopezcampos@qub.ac.uk.

of individual genetic variation is an aspect that has not been explored in detail in the area of toxicogenomics. Therefore, in this work we propose the development of a computational analytical pipeline designed to integrate the increasing amounts of information related with human genetic variation and the chemical component of the exposome analyzing the effects of different coding variants in the interactions with chemical compounds.

2. Methods and Results

As a proof of concept for our approach we have chosen to analyse the effects of the different coding genetic variants in the human Estrogen Receptor 1 (ESR1) and how those would affect the interactions with bisphenol-A (BPA), a chemical compound known to link with ESR1 and a potential endocrine disruptor due to its actions as a xeno-estrogen.

For the analyses we selected the wild type (WT) sequence and five single nucleotide variants (SNV) (R269C, R269H, M315V, E330K and R477Q) in the ESR1 gene. These variants were chosen as they were condign variants, they were not known to be pathogenic, were not located in the catalytic/binding site of the protein and had some of the higher minor allele frequencies for the variants in this gene.

We first used I-Tasser [2] to predict the ESR1 structure beyond the resolved region available and then we modelled our five variants using SCWRL [3]. Finally, we used those models to simulate the protein/ligand docking and calculate the binding affinities for BPA and estradiol for each of the five mutants and the wild type structures using CANDOCK [4]. As expected the results showed (table 1) that there was a higher affinity in this case for the natural ligand (estradiol) than for the xenobiotic (BPA) as well as that the effect in the reduction of the binding activity was more intense for the natural ligand than for BPA in all the analysed variants.

Table 1. Table showing the binding affinities (Kcal) for the wild type and the five variants analysed.

	Wild type	R269C	R269H	M315V	E330K	R477Q
Estradiol	-64.2911	-48.8941	-48.8198	-46.3230	-50.8965	-47.6796
BPA	-47.4722	-36.5059	-39.6272	-41.1594	-39.5739	-41.2424

With this approach we brought together genomics and exposomics applying in silico approaches. This strategy can be extended for the analyses of other targets and variants for the different chemicals to which individuals are exposed daily

References

[1] C.P. Wild. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev.* **8** (2005), 1847-50

[2] A. Roy, A. Kucukural, Y. Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, **5** (2010), 725-738

[3] Q. Wang, A.A. Canutescu, R.L. Jr. Dunbrack. SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. *Nat Protoc.***12** (2008), 1832-47

[4] J.A. Fine, J. Konc, R. Samudrala, G Chopra. CANDOCK: Chemical atomic network based hierarchical flexible docking algorithm using generalized statistical potentials. *BioRxiv*,(2018) 442897

Lower Respiratory Tract Infections in Children, Weather and Google Search Trends

Manuel RODRÍGUEZ TABLADO^a, Paula GONZALEZ PANNIA^a, Fernando C. FERRERO^a, Santiago ESTEBAN^a

^a*Health Ministry, Ciudad Autónoma de Buenos Aires, Argentina.*

Abstract. This is an exploratory study on different variables known to be associated with epidemiological outbreaks of lower respiratory tract infections in children under 2 years raised to identify patterns and potential predictors for making strategic decisions in terms of strengthening health teams in periods of greatest demand.

Keywords. Electronic Health Record, lower respiratory tract infection, weather, Google search.

1. Introduction

Lower respiratory tract infections (LRTI) are one of the most frequent causes of medical consultation, hospital admission and mortality in children with a very marked seasonal incidence [1]. These characteristics determine healthcare systems to be overloaded at the times of the year with higher demand for this group of pathologies.

This is an exploratory study on different variables known to be associated with epidemiological outbreaks of LRTI in children under 2 years [2], [3] raised to identify patterns and potential predictors for making strategic decisions in terms of strengthening health teams in periods of greatest demand.

2. Methodology

We analyzed the proportion of LRTI consultations and hospital admissions for all causes in patients under 2 years, climatological variables (air temperature, humidity and atmospheric pressure) and Google search trends on respiratory symptoms keywords. Data of consultations were extracted from the EHR of the public health system of the City of Buenos Aires. Consultations related to LRTI were detected using a rule-based algorithm validated using as gold standard the manual review of the medical records (precision 85.96%, recall 82.35%, F1 score 84.12%). Hospital admissions data were extracted from an administrative database of all the hospitals of the public health network of the City of Buenos Aires and we included admissions for all causes due to many missing data about the cause of admission (35.5% of missing data). The information related to Google searches included the terms "cough", "bronchiolitis" and "flu" in Spanish (pediatrician experts selected them as main terms used by parents as cause of consultation) and the average relative popularity of these terms was calculated. We

analyzed weekly averages or accumulated totals since 2017-01-01 to 09-30-2019 to assess seasonal trends and the relationships between them.

3. Results

We found a very strong relationship between air temperature (Spearman's rank correlation coefficient -0.89 , $p < 0.001$) and Google search trends (ρ 0.89 , $p < 0.001$) with the proportion of medical consultations for LRTI (**figure 1**). It is interesting that the LRTI consultation curve rises rapidly in the first weeks in which the temperature begins to fall. Google searches do not seem to have a pattern that allows for predictive analysis. Weaker relationship was found with hospital admissions (ρ 0.44 , $p < 0.001$), atmospheric pressure (ρ 0.62 , $p < 0.001$) and humidity (ρ 0.40 , $p < 0.001$).

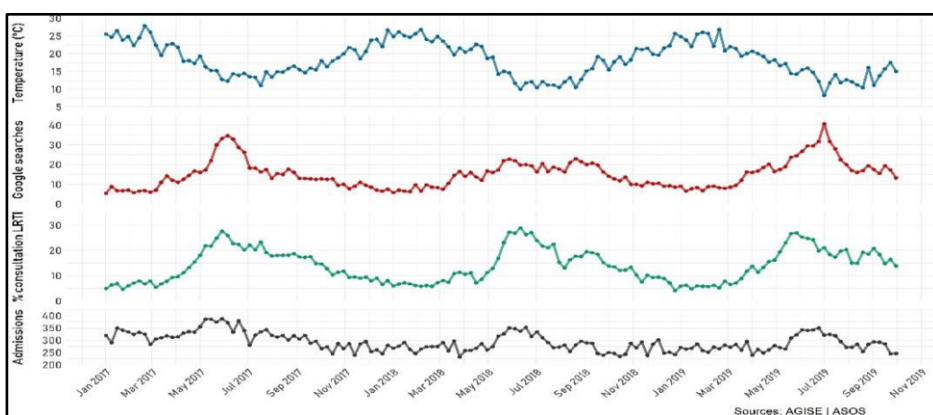


Figure 1. Relationship between air temperature and Google search trends with the proportion of medical consultations for LRTI.

4. Conclusions

This is a novel study about climatological variables and clinical information using the EHR in the City of Buenos Aires. Further investigation should be made in order to make informed decisions regarding strategies to strengthen the health care systems in times of greatest demand. Air temperature seems to be the most important factor that could predict new outbreaks of LRTI.

References

- [1] C. Troeger *et al.*, "Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *Lancet Infect. Dis.*, vol. 18, no. 11, pp. 1191–1210, Nov. 2018.
- [2] S. Pollett *et al.*, "Evaluating Google flu trends in Latin america: important lessons for the next phase of digital disease detection," *Clin. Infect. Dis.*, vol. 64, no. 1, pp. 34–41, 2017.
- [3] C. Wang *et al.*, "A randomized clinical trial of exercise during pregnancy to prevent gestational diabetes mellitus and improve pregnancy outcome in overweight and obese pregnant women.," *Am. J. Obstet. Gynecol.*, vol. 216, no. 4, pp. 340–351, Apr. 2017.

Manual Evaluation of the Automatic Mapping of International Classification of Diseases (ICD)-11 (in French)

Julien GROSJEAN^{a,b}, Kévin BILLEY^{a,c}, Jean CHARLET^{b,d}, Stefan J. DARMON^{a,b}

^aDepartment of Biomedical Informatics, Rouen University Hospital, Normandy, France

^bLIMICS U1142 INSERM, Sorbonne Université, Paris & Rouen University, France

^cLITIS EA4108, Rouen University, France

^dAssistance Publique-Hôpitaux de Paris, DRCI, Paris, France

Abstract. A lexical method was used to map ICD-11 to the terminologies included in the HeTOP server. About half of ICD-11 codes (47.76%) were mapped to at least one concept. The developed tool reached a global precision of 0.98 and a recall of 0.66. Lexical methods are powerful methods to map health terminologies. Supervised and manual mapping is still necessary to complete the mapping.

Keywords. Terminology mapping; Semantics; Terminology as topic; Unified Medical Language System; Vocabulary, controlled

1. Introduction

The International Classification of Diseases (ICD) version 11¹ is a major revision of the ICD-10 which is widely used to code diagnoses in healthcare systems. Furthermore, the purpose of the Unified Medical Language System® (UMLS) [1] is to facilitate the development of computer systems that behave as if they "understand" the meaning of the health language. The process of terminology mapping consists of identifying identical (or approximately identical) concepts or relationships between terminologies [2]. Since 2007, the Rouen University Hospital (RUH) Department of Biomedical Informatics (DBI) has developed the HeTOP cross-lingual health termino-ontology server [3], which includes 75 terminologies (or ontologies) in 32 languages².

The objective of this work is to propose a mapping of ICD-11 to the UMLS Metathesaurus and to French terminologies which are not (yet) included in the UMLS.

2. Methods

French natural language processing tools and mapping algorithms were developed by the RUH DBI team to map health terminologies in French. These tools were used in previous works [3][5] and extended to link terms in multiple French health terminologies. This approach allows from a given term, to find a HeTOP concept with French (or English) terms that are most lexically similar to it.

¹WHO International Classification of Diseases, 11th Revision (ICD-11); <https://www.who.int/classifications/icd/en/> (Accessed September 5, 2019).

² <https://www.hetop.eu/>

3. Results

The RUH DBI automatic mapping tool has generated 144,823 ICD-11 mappings to the 75 terminologies integrated into HeTOP. Concerning mappings between ICD-11 and ICD-10, WHO has manually provided 12,767, whereas RUH DBI has provided 6,925. The intersection between WHO and RUH DBI is 3,492 mappings (respectively 27.35% and 50.43%). This French automatic mapping tool has provided at least one mapping for 28,521 ICD-11 distinct codes (using their preferred labels and its synonyms and acronyms) among 59,709 ICD-11 codes (47.76%). The Top 6 terminologies with the most mappings to ICD-11 are SNOMED CT, MeSH and NCIT, which is coherent with the size of these terminologies and their translations as well as their main objectives, respectively 326,927 (201,659 in French), 378,024 (including the MeSH concepts; 105,933 in French) and 148,215 (79,409 in French). Among the 144,823 automatic mappings from ICD-11 to the 75 terminologies integrated into HeTOP, 2,371 were manually supervised by one expert (SJD), using six relations types: EM (exact match), BTNT (broader than), NTBT (narrower than), RT (related), unknown and false. The total number of distinct ICD-11 codes manually curated was 464 (mean mappings number of 5.1 per code). 82% of these automatic mappings were validated in exact match. 15% were evaluated as “close” mappings (i.e. BTNT, NTBT, RT). Only 46 mappings (2%) were evaluated as false. Furthermore, among the 464 ICD-11 curated codes, 183 have no mappings to any HeTOP concept. On the other hand, the expert manually added new mappings to 26 distinct ICD-11 codes without any automatic mappings. Thus the false negative rate is estimated to $26/(183+26)=12.4\%$. 1,226 new mappings from transitive closure were also added whose 154 automatically validated as Exact Match. Those figures allow an estimation of the recall: $2,371/(2,371+1,226)=65.92\%$.

4. Discussion and conclusion

To the best of our knowledge, this is the first study to map the entire ICD-11 classification to terminologies in French. The quality of these mappings can be considered as very good, as 82% were evaluated as an “exact match”, and only 2% are false. However, the cover of such tool is relatively low (47.76% of all ICD-11 codes) as many labels are too precise or not based on a state-of-the-art terminological method.

References

- [1] D.A.B. Lindberg, B.L. Humphreys, A.T. McCray. The unified medical language system. *Methods Inf Med* **32** (1993), 281–91.
- [2] A.R. Aronson. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. *Proc AMLA Symp* (2007), 17–21.
- [3] J. Grosjean, T. Merabti, B. Dahamna, I. Kergourlay, B. Thirion, L.F. Soualmia, S.J. Darmoni. Health multi-terminology portal: a semantic added-value for patient safety. *Stud Health Technol Inform*. **166** (2011), 129–38.

Master Patient Index Standardization Patient Search Identification Service (PSIS) of the National Directorate of Health Information Systems (DNSIS) Argentina

Humberto F MANDIROLA BRIEUX^{a, b} Alejandro LOPEZ OSORNIO^a Martin DIAZ
MAFFINI ^a Miguel AMORESE^a and Daniel A RIZZATO LEDE ^a

^a Ministry of Health, Argentina

^b HL7 Argentina.

Abstract. Correct patient identification is the cornerstone for the proper implementation of electronic health records. Up to 20% of the registered patients are duplicated in most systems. Strong identification policies and robust systems can minimize such errors. In this poster we share the Ministry of Health recommendations for the Master Patient Index improvement using search algorithms.

Keywords. Master Patient Index, Electronic Health Records, Soundex, Levenshtein, Patient Search Identification Service

1. Introduction

One of the most serious problems for Master Patient Index (MPI) is duplication of data[1]. Conventional database systems generally use deterministic searches to find document numbers or patient names, leading to mistakes. When the person responsible for entering data in the MPI is unable to find the patient, she will reenter the information, thus creating a duplicate patient. Duplicated patients in the MPI poses the risk of the loss of data due to the fragmentation of information [4]. It is important to use Patient Search Identification Service (PSIS) to avoid these types of mistakes. By using algorithms for a probabilistic search with multiple data (name, surname, birthdate, id type and number, gender), it is possible to identify suitable candidate patients by a group of similar data[2]. PSIS allows us to identify potential similar candidates and provides significant advantages over conventional searches [3] keeping the MPI clean.

2. Materials and Methods

For our Model of PSIS we use a minimum set of patient data composed of six data points that are: **Surname** (first and following), **Name** (first and middle), **Gender**, **Document Type**, **Document Number** and **Birthdate**. We apply two combined algorithms, the first being phonetic (Soundex), the other comparative (Levenshtein).

PSIS makes the first search for candidates by using Soundex and the second level by the distance of Levenshtein’s algorithm and that allows us to find the patients candidates files, see figure 1

The variables that consider are the relative weight V, the distance of Levenshtein and the length of the text of the candidate L. The coefficient F is calculated with the formula $F = (L-D) / L$ and finally, the value of each variable is obtained by multiplying the weight V by the coefficient F. The sum of the 6 variables gives us the probability that the candidate is the wanted one, in this case, 73.75%. See figure 2

3. Results

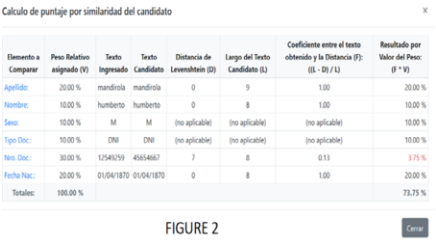
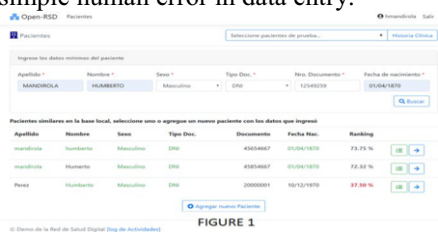
In order to prove the advantages of the PSIS we compared the Traditional Database Search (TDS) with PSIS. Results showed that the PSIS we built was a substantial help. PSIS allows us to detect 350 duplications in an MPI that had 50.000 patients, compared to TDS detecting only 50 duplications.

4. Discussion

The PSIS significantly improves the discovery of duplicates in an MPI compared to that of TDS. PSIS searches identify 88% more duplicate candidates than those found in TDS.

5. Conclusion

If we make a mistake in the document number by a single digit, using conventional searches (TDS), the patients cannot be found. Using the Levenshtein distance algorithm allows us to identify that the number is very similar to the one searched often indicating a simple human error in data entry.



References

[1] Just BH, Marc D, Munns M, Sandefer R. Why Patient Matching Is a Challenge: Research on Master Patient Index (MPI) Data Discrepancies in Key Identifying Fields. Perspect Health Inf Manag [Internet]. 1 de abril de 2016 [citado 23 de diciembre de 2019];13(Spring). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4832129/>

[2] Deo RC. Machine Learning in Medicine. Circulation. 17 de noviembre de 2015;132(20):1920-30.

[3] Morris G, Afzal S, Finney D, Farnum G, Robinson C, Greene J, et al. Patient Identification and Matching Final Report. 2014;93.

[4] Patient and Sample Identification. Out of the Maze? [Internet]. [citado 24 de diciembre de 2019]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5471642/>

Near Real Time Feedback of Seasonal Influenza Vaccination and Virological Sampling: Dashboard Utilisation in a Primary Care Sentinel Network

Harshana LIYANAGE^a, Oluwafunmi AKINYEMI^a, Sameera PATHIRANNAHELAGE^a, Mark JOY^a and Simon DE LUSIGNAN^{a,b,1}

^a*Nuffield Department of Primary Health Care Sciences, University of Oxford, UK*

^b*Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC), London, UK*

Abstract. Near-real time feedback on data quality allows clinicians and data managers at general practices to collectively contribute to improving the reliability of observational data. We report features of an enhanced dashboard used for virological surveillance after the first season of utilisation. Time-series analysis performed on dashboard usage data from 68 general practices (denominator = 722,409) indicated utilisation patterns that correlated to influenza vaccination in general practices. Near real time feedback via a dashboard may improve data quality and possibly vaccine uptake.

1. Introduction

Influenza contributes significantly to the burden of disease morbidity and mortality worldwide[1]. Virological surveillance helps identify the predominant strain in circulation; it also provides an understanding of the pattern of disease transmission and planning towards pandemic preparedness. We report enhancements made to the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) virological surveillance dashboard and usage analysis after its first use in surveillance activities during a complete influenza season (2018/2019) [2].

2. Methods

We obtained data for the dashboard from the RCGP RSC database which includes pseudonymised health care records of individuals registered with general practice network in England (68 practices that participated in the virology swabbing scheme with an overall denominator of 722,409) [3]. These practices provided nasal and nasopharyngeal swabs for confirming influenza within a laboratory setting. We enhanced the dashboard for the 2018/2019 influenza season by adding comparative assessment

¹ Corresponding Author: Simon de Lusignan, Nuffield Department of Primary Health Care Sciences, University of Oxford, UK; E-mail: simon.delusignan@phc.ox.ac.uk.

episode type recording customised for each given practice. We carried out time-series modelling on the dashboard data (from July 2018 to March 2019) using the R statistical package (version 3.6.1) and compared with seasonal influenza and flu vaccination trends.

3. Results

The participating general practices accessed the dashboard particularly between October 2019 and March 2020 which is the peak of the flu season. General practices typically start conducting flu clinics at the beginning of October (in any given year) and are held weekly for about 8 weeks. We observed a peak in dashboard access following the initial flu vaccination campaign and a 2nd peak in dashboard access soon after the increase of Influenza-like (ILI) incidence at the beginning of January 2019. Results of the Granger causality test indicated that there is evidence that after controlling for the lagged effects of flu vaccination counts, lagged visits carry predictive power ($p < 0.001$) for flu vaccinations. There is no evidence that the same applies between ILI counts and visits.

4. Discussion

Trend analysis may explain the periods when dashboard utilisation increases across the sentinel network. The results indicate that dashboard use increases shortly after influenza vaccination campaigns with the peak seen during influenza circulation not being statistically associated.

5. Conclusion

Findings of this study together with feedback from user surveys helps the RCGP RSC to understand the peaks of dashboard use, and which areas of the dashboard have more user engagement that could benefit from further development. Dashboard feedback may have a role in improving data quality and prompting vaccine administration.

References

- [1] L.C. Ramsay et al., S.A. Buchan, R.G. Stirling, B.J. Cowling, S. Feng, J.C. Kwong, and B.F. Warshawsky, The impact of repeated vaccination on influenza vaccine effectiveness: a systematic review and meta-analysis, *BMC Med.* **17** (2019) 9.
- [2] S. Pathirannehelage et al., Uptake of a Dashboard Designed to Give Realtime Feedback to a Sentinel Network About Key Data Required for Influenza Vaccine Effectiveness Studies, *Stud Heal. Technol Inf.* **247** (2018) 161–165.
- [3] A. Correa et al., Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile, *BMJ Open* **6** (2016) e011092.

Neural Networks for Cause of Hospitalization and Final Cause of Death Extraction from Discharge Summaries

Francisco E. PEPPER^{a,1}, M. Belén CARDONE^b and Santiago ESTEBAN^a

^a*Family and Community Medicine Division, Hospital Italiano de Buenos Aires, Argentina*

^b*Instituto Universitario Hospital Italiano de Buenos Aires, Argentina*

Abstract. Determining the cause of death of hospitalized patients with cardiovascular disease is of the utmost importance. This is usually recorded in free text form. In this study we aimed to develop a series of supervised natural language processing algorithms to extract cardiovascular causes of hospitalization and final causes of death.

Keywords. Natural language processing, discharge summaries

1. Introduction

Determining the cause of death of hospitalized patients with cardiovascular disease is of the utmost importance in order to take measures and thus improve the quality of care of these patients and prevent preventable deaths. Even though hospitalization causes and final causes of death can be recorded in structured form within electronic medical records (EMR), they are usually entered in free text form inside discharge summaries[1]. In this study we aimed to develop a series of supervised natural language processing algorithms in order to extract cardiovascular causes of hospitalization and final causes of death.

2. Methods

We manually reviewed a random sample of 6161 discharge summaries of patients who died between January 1st, 2005 and December 31st, 2016 at Hospital Italiano de Buenos Aires, Argentina. All annotations were carried out by one family physician. Patients' cause of hospitalization was initially classified as cardiovascular causes vs non-cardiovascular. Cardiovascular causes were further classified into ICD-10 codes. These comprised 21.3% (n = 1316) of the total sample. We defined three classification tasks: two related to cause of hospitalization (cardiovascular vs non-cardiovascular; specific type of cardiovascular cause) and one related to final cause of death. Causes of hospitalization were annotated as cardiovascular vs non-cardiovascular and then as ischemic cardiac diseases, cerebrovascular diseases and other causes. Final causes of death were classified in six categories namely, cardiogenic/hypovolemic shock, septic shock, cerebrovascular diseases, renal failure, malignant neoplasms and other causes. The sample was randomly split into training (80%), validation (10%) and test sets (10%). We then trained three different algorithms for the three classification tasks and evaluated their performance on the test set. For the three algorithms different architectures were

¹ Corresponding author email francisco.peper@hospitalitaliano.org.ar

tested (Support vector machines, logistic regression, k-nearest neighbors, XGBoost, Random Forests, C5.0, GBM and different neural network architectures. For the three cases, neural networks showed the best results in the validation set. The final architecture consisted of a bi-directional long-short-term-memory recurrent neural network with attention mechanism, with two stacked dense layers. Words were represented using FastText word vectors.

3. Results

For the first classification task (algorithm 1), the performance on the test set resulted in an accuracy of 0.9546 (95% CI 0.9351, 0.9696), Recall of 0.979, Precision of 0.9225 and F1-score of 0.895. Table 1 shows the individual metrics per category for the second and third algorithms.

Table 1. Performance of the second and third algorithms on the test set.

Algorithm 2						
		Ischemic cardiac diseases (I20-25)	Cerebrovascular diseases (I60-I69)		Other cardiovascular causes	
Recall		0.67	0.967		1	
Precision		1	1		0.9	
F1-score		0.815	0.96		0.95	
Algorithm 3						
	Cardiogenic hypovolemic shock (R57.0/R57.1)	Septic shock (R57.2)	Cerebrovascular diseases (I60 - I69)	Renal failure (N17-N19)	Malignant neoplasms (C00-C97)	Other causes
Recall	0.94	0.47	0.76	0	-	0
Precision	0.79	1	0.59	-	-	-
F1-score	0.86	0.64	0.66	-	-	-

4. Discussion

Even in this low sample size scenario, neural networks outperformed classical machine learning algorithms for text classification. Nevertheless, performance dropped significantly as the sample size was reduce for algorithms 2 and 3 and also as the number of classification categories increased.

5. Conclusions

We developed three NLP algorithms for information extraction regarding cause of hospitalization and final cause of death from discharge summaries.

References

[1] Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: A literature review. *J Biomed Inform.* 2018;77:34–49. doi:10.1016/j.jbi.2017.11.011

Numerical Analysis of Conventional Air Pollutants and PM₁₀ Concentration Affected to Respiratory Disease Patients in Bangkok, Thailand

Wisit THONGKUM^{a,1} and Santisith KHIEWKHERN^a

^a*Faculty of Public Health, Mahasarakham University, THAILAND*

Abstract. Ambient air pollution is a major cause of death and disease globally. The linear regression between respiratory disease patients, PM₁₀ and other criteria pollutants were used in this study. Among respiratory disease patients and CO was statistically significant with P-value of <0.01 ($r=0.198$, 95% CI=48.74 - 288.16). As for respiratory disease patients and NO_x was statistically significant with P-value of <0.01 ($r=0.190$, 95% CI=1379.74 – 9241.43).

Keywords. Air pollutants, PM10, Respiratory disease

1. Introduction

Air pollution kills an estimated 7 million people worldwide every year. Air pollution resulting from traffic is a major problem in many cities in Asia, including Bangkok, Thailand. This pollution originates mainly from incomplete fossil fuel combustion, transportation, and construction (Ruchirawat et al, 2007). The concept of this study is utilizing the data obtained from the current conventional air monitoring station in Bangkok to analyze conventional air pollutants and PM₁₀ concentrations affected respiratory disease patients in this area.

2. Data analysis and Method

2.1. Data collection and Analysis

The air monitoring data used in this study were from Jan 2013 to Dec 2018. Conventional air pollutants and PM₁₀ were monitored at 3 monitoring sites in The Bangkok Metropolitan area. The data of respiratory disease patients in this study from Jan 2013 to Dec 2018. The respiratory disease was bronchitis, emphysema, chronic obstructive pulmonary disease, and asthma. This study was collected from 3 hospitals. The correlation coefficient and linear regression analysis were used to determine a potential correlation between conventional air pollutants and PM₁₀ concentrations and number of respiratory diseases patients. Independent variables were conventional air pollutants and PM₁₀ concentrations. The dependent variable was respiratory disease patients.

¹ Corresponding Author; E-mail: santisith.k@msu.ac.th.

3. Results and Discussions

3.1. Diurnal variation of PM₁₀ concentration and respiratory disease patients

The concentration of PM₁₀ shows a similar type of variation as a decrease from May to Sep an increase from Oct to Dec and Jan to Apr of each year. This is probably because of the enhance emission from traffic and seasonal variation (Aurangojeb M., 2011). The number of respiratory disease patient’s tendencies increase from Jan 2013 to Dec 2018. The number of respiratory disease patient’s tendency decrease from Jan 2013 to Apr 2017 after that increase to Dec 2018. The concentration of PM₁₀ shows a similar type of variation as a decrease from May to Sep an increase from Oct to Dec and Jan to Apr of each year. The number of respiratory disease patients slightly varies from Jan 2013 to Dec 2018. After the government of Thailand has taken various roles to mobilize its resources for haze management in the second stage from 2011 to 2019.

3.2. The correlation coefficient and linear regression analysis

The Pearson’s product moment correlation coefficient (r) was used to determine the strength of association between a pair of variables, to test the relationship between these variables and to test whether the association is greater than could be expected by chance (Rummasak T., 2019). Among respiratory disease patients and CO was statistically significant with P-value of <0.01 (r=0.198, 95% CI=48.74 - 288.16). As for respiratory disease patients and NO_x was statistically significant with P-value of <0.01 (r=0.190, 95% CI=1379.74 – 9241.43) presented in Table 1.

Table 1. Linear regression analysis of respiratory disease patients, PM₁₀ and other criteria pollutants

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
2 (Constant)	366.33	110.14		3.326	.001	149.22	583.45
CO	168.45	60.73	.198	2.774	.006	48.74	288.16
NO _x	5310.59	1994.17	.190	2.663	.008	1379.74	9241.43

4. Conclusions

The concentration of PM₁₀ shows a decrease from May to Sep an increase from Oct to Dec and Jan to Apr of each year. The number of respiratory disease patient’s tendency decrease from Jan 2013 to Apr 2017 after that increase to Dec 2018. The linear regression between respiratory disease patients, PM₁₀ and other criteria pollutants were used in this study. Among respiratory disease patients and CO, respiratory disease patients and NO_x were statistically significant.

References

[1] Aurangojeb M., Relationship between PM10, NO2 and particle number concentration: validity of air quality controls, Procedia Environmental Sciences 6 (2011), 60-69.

[2] Ruchirawat M., Settachan d., Navasumrit P., Tuntawiroon J., and Autrup H. Assessment of potential cancer risk in children exposed to urban air pollution in Bangkok, Thailand, Toxicology Letters 168 (2007), 200-209.

[3] Rummasak T. Relationship between annual average concentration of ambient PM10 and out-patients with respiratory disease: Thailand case study, Applies Environmental Research 41(1) (2019), 57-69.

Patients' Experience and Assessment on the Transition from Paper to Electronic Medical Records

Carolina ARMAGNAGUE THIVANT^a, Santiago ESTEBAN^a
Analia J. BAUM^a

^a*Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina.*

Abstract. By the implementation of the EMR in the Primary Care Centers of the City of Buenos Aires, it has been claimed that there is a need to obtain systematic and accurate information about the assessment that patients have about such incorporation. A qualitative research has been carried out in order to answer this query.

Keywords. Doctor-patient relationships; Electronic medical records; paper health record.

Introduction

The *Health Information Systems Office* at the Buenos Aires City health ministry EMR launched the implementation of the EMR in the First level of Health Care in 2016. This comprises 44 primary care health centers (PCHC). In this context, we have recognized the need to obtain systematized and accurate information about patients' experience during the transition from the paper health record (PHR) to the EMR.

Methods

We conducted a qualitative exploratory study. We used semi-structured interviews of a convenience sample of patients. The participants selected were men and women over 18 years of age who at least once had been treated in a center where the EMR had been implemented. Three PCHC were selected as these were the first to implement the EMR. Interviews were audio-recorded and transcribed verbatim. The dimensions were extracted into a database table and categories were created depending on similarity and repetition among relevant segments to approach the problem.

Main outcomes

We interviewed 29 patients. The sample was divided based on age: "Young" (18<30); "Adult" (30<60); "Older Adult" (60<). Most of the interviewees considered that the

EMR is the same or better than the PHR. In relation to EMR's attributes, they highlighted: practicality, exhaustiveness, speed, and the possibility of being in the network. As to the effects of the EMR implementation on the medical visit, they considered that it represents a substantial help to the professional. It also avoids repetition of information and allows for focusing directly on the reason for the medical visit, providing an integral view of the patient. With respect to the PHR, interviewees considered that the process involved in the PHR was more tedious or slower. It can be observed that the EMR prevails over the PHR. Nevertheless, patients in the three age groups insisted on the importance of having an updated backup copy in paper. Patients had a positive opinion towards health professionals' advanced use and skills of the EMR. They believe that health professionals are satisfied with the EMR implementation and that the time required to register information into the EMR is reasonable. As for the young and the adult groups, it can be observed that there is an interest in sharing the screen with the professional in order to help register the information. Instead, the older adults' group stated that they do not find it necessary to see the screen as they "trust" in the professional. As emerging topics, the interest of young people in the possibility of reprinting lost prescriptions or vaccination certificates appeared; at the same time, they showed interest in having access to the EMR at home.

Discussion

Most of the interviewees viewed that the EMR is the same as the PHR while others considered that the former is better. This result led us to infer that the EMR implementation has not been experienced as an interruption or breakdown in the attention, but instead as a continuity of their health care. As can be seen from the results, patients consider that professionals have a positive assessment on the EMR and that they master this new system. As other studies reveal, the professional's positive attitude towards the use of the EMR can impact on the patient's assessment and experience [1].

The sample was divided based on age because the team considered that differentiation could directly influence on the naturalization of digital devices to daily life. In effect, the results obtained in relation to the young people would lead to think in new user generations of the health system with an active position and willingness to gain more participation. Unlike that perspective, within older adults, it prevails the idea that the doctor represents a figure of power and authority that hinders the possibility of thinking in a method of cooperative and symmetric attention.

Conclusion

Patients generally express a positive or neutral perception of the EMR. There are not elements to infer a loss of continuity of care.

References

- [1] Lelievre S, Schultz K. Does computer use in patient-physician encounters influence patient satisfaction? Canadian family physician *Medecin de famille canadien* 56 (2010), 6-12.

Personal Health Record (PHR) System in Portable Health Clinic

Rafiqul ISLAM^a, Kimiyo KIKUCHI^b, Yoko SATO^c, Rieko IZUKURA^a,
Fumihiko YOKOTA^b, Mariko NISHIKITANI^b, Raisa TASNEEM^d,
Nazneen SULTANA^d, Ashir AHMED^e and Naoki NAKASHIMA^a

^aMedical Information Center, Kyushu University Hospital, Japan

^bInstitute of Decision Science for Sustainable Society, Kyushu University, Japan

^cDepartment of Medical Science, Kyushu University Hospital, Japan

^dGrameen Communications, Dhaka, Bangladesh

^eDepartment of Advanced Information Technology, Kyushu University, Japan

Abstract. Personal Health Record (PHR) is not just the collection of personal health data but also a personal healthcare and disease management tool for the individual patient as well as a communication tool with the medical staff. Moreover, recently PHR has been considered an indispensable tool for patient engagement in the area of non-communicable diseases (NCDs) and has gained importance. Like many other developing countries, the growth of NCDs is very high in Bangladesh. Portable Health Clinic (PHC) system has been developed there with a focus on NCDs and PHR is there from the beginning. This study for the standardization of PHR system of PHC with the reference of the PHR proposed by Japanese Clinical Societies could be a reference work for the national PHR system development in the country.

Keywords. Personal Health Record, Portable Health Clinic, Standardization

1. Introduction

Personal Health Record (PHR) has been getting popular as an important supportive tool for providing better healthcare service by the medical professionals. These days there are many applications for smartphones that help to manage personal clinical data about NCDs generated in clinics and drugs dispensed by pharmacies. This is the simplest form of PHR. However, a standard PHR system is expected to manage information regarding an individual's vital signs, physical exercise, behavior, and location at home or at the office using built-in sensors in the smartphone or using sensor networks based on internet of things (IoT) technology. Patient will have access to monitor this personal health/medical data and express personal opinion for the progress of the treatment. This helps the physician to be confident to change the policy and to intervene [1].

2. Portable Health Clinic

Portable Health Clinic (PHC) system was designed in 2010 by Grameen Communications, Bangladesh in a joint collaboration with Kyushu University of Japan as a telemedicine

based healthcare system [Figure 1]. Later some new modules have been added to this modular system for satisfying patients demand for serving secondary level of treatments in Eye (ophthalmic) Care and Maternal & Child Healthcare (MCH) [2].

PHR service has been there as a personal service to the patients from the beginning. The patients can access to their own data on smartphone, monitor and share their feedbacks to the physicians if they need to receive any consultancy.

3. Standardization of PHR

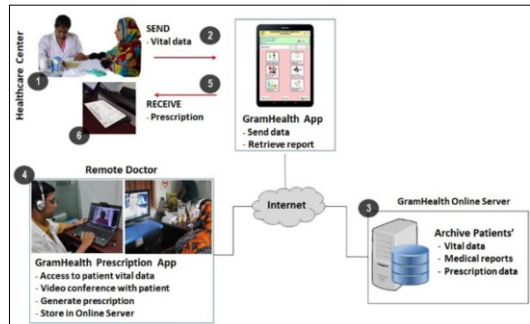


Figure 1 Work Flow of PHC System

As the content of the PHR of PHC system was developed independently, there are always some issues for the use of this PHR data by medical professionals of other countries in terms of clinical data items and measurement units. The clinical data items in PHR varies country to country. Thus, the need of standardization of the PHR has always been felt for data portability and to pass interoperability.

Recently six major Japanese clinical societies jointly proposed a set of 41 items as standard for PHR for NCDs in Japan after years of investigations [3]. There was unanimous consensus, then installed and expanded allover Japan. This PHR protocol is being used in various top PHR service provider (app company) in Japan.

This research aimed to standardize PHC's PHR system considering this Japanese PHR system as a standard and passed the interoperability for NCDs. In case of Eye Care, the ophthalmic image in JPEG format has been added in PHR for diseases prediction by AI based image analysis. In case of MCH, hemoglobin measurement and fetus beat per minute has been added as these are the major issues for mother and child respectively.

4. Conclusion

Development of PHR system and its standardization is very much desired for wider scope of quality healthcare services in Bangladesh from medical service point of view. This work of PHR standardization of PHC system is just a step forward in this direction.

References

- [1] Y Bombard, GR Baker, E Orlando, et al. Engaging patients to improve quality of care: a systematic review. *Implement Sci.* 2018;13(1):98. Published 2018 Jul 26. doi: 10.1186/s13012-018-0784-zTechnology and Tools for Patient Engagement
- [2] R Islam, Y Nohara, MJ Rahman, N Sultana, A Ahmed, N Nakashima, "Portable Health Clinic: An Advanced Tele-Healthcare System for Unreached Communities", *Ebook: MEDINFO 2019: Health & Wellbeing e-Network for All*, Volume 264, 416-419, Aug, 2019.
- [3] N Nakashima, M Noda, K Ueki, et al. Recommended configuration for personal health records by standardized data item sets for diabetes mellitus and associated chronic diseases: A report from Collaborative Initiative by six Japanese Associations. *Diabetol Int.* 2019;10:85-92 <https://doi.org/10.1007/s13340-019-00389-7>

Persuasive eHealth to Support Home Rehabilitation of the Elderly After a Hip Operation: An Explorative Approach

Laurence ALPAY^{a,1} and Ybranda KOSTER^a

^a Medical Technology Group, Inholland University of Applied Sciences

Abstract. Home rehabilitation after a hip operation can be daunting for the elderly. Lack of motivation to exercise and being insecure in the recovery process are common barriers. Personalized eHealth can help to ensure that the patient exercise efficiently, filling the gap between treatment in the practice with the physical therapist and practice at home.

Keywords. eHealth, home rehabilitation, elderly

1. Introduction

The elderly population is growing worldwide, and osteoarthritis amongst this group is common, often resulting in a joint prosthesis¹. The stay in the hospital after such an operation is very short. Rehabilitation after discharge from the hospital takes place with a physical therapist. The usual after-care is about exercising to improve the mobility of the hip / knee. Instructions are usually given in the form of folders of videos. However, physical therapists see a gap between treatment in their practice and practice at home. Furthermore, patients are not always motivated enough to exercise at home and are unsecure in what they do. The use of eHealth to support the patient's rehabilitation at home is limited. Sensor technologies for example are used to collect quantitative information about joint recovery². The research reported here focuses on exploring the design of a persuasive eHealth solution, for effective rehabilitation at home.

2. Method

In order to design a persuasive eHealth solution, we use the methodology we developed based on the CeHRes framework and other models for persuasive design and behavioral change³. This methodology begins with a contextual inquiry, resulting into prototyping and finally implementation. Persuasive features and triggers for motivational behavior are throughout the design process incorporated into the eHealth solution. Three groups of 3rd and 4th year bachelor students (4 to 5 students per group) worked for a period of 20 weeks with the Medical Technology Research Group on this project. One group

¹ Corresponding Author: Laurence Alpay, Inholland University of Applied Sciences, Bijdorplan 15, 2015 CE Haarlem, The Netherlands; E-mail: Laurence.alpay@inholland.nl

consisted of informatics students from InHolland University of Applied Sciences and two other groups consisted of a multidisciplinary team from Rotterdam University. Students used a fictive casus reflecting a real-life situation. We chose the approach of student's participation and action research in order to explore various proofs of concept and to provide students the opportunity to experience a healthcare setting and to think out of the box with other disciplines.

3. Results

The designing of an eHealth solution focused mainly contextual inquiry, requirements and prototyping. Formative evaluation was carried out with the targeted groups. Each group provided a proof of concept developed in co-creation with the targeted group and physical therapists. Researchers from the Medical Group acted as eHealth experts. Figure 1 shows a paper prototyping from one group during contextual inquiry, and a developed solution to ensure that the exercises are done correctly from another group. Students used Fogg's Behavior Grid to determine the kind of behavior change that was needed to optimize the rehabilitation process at home. In the design of their eHealth solutions they used the principles of persuasive design as described by Oinas Kukkonen to determine which persuasive elements should be used.



Figure 1. Prototype examples

4. Discussion and Conclusions

Home rehabilitation after a hip or knee operation often remains a challenge for a growing number of elderly and limit participation and self-management. Exploring various eHealth proof of concepts based on persuasive features and behavioral change can help to tune to a more personalized solution. Students' participation is an added-value for the healthcare setting and for providing input to further research in this area.

References

- [1] D.J. Healee et al, Older adult's recovery from hip fracture: A literature review, *International Journal of Orthopaedic and Trauma Nursing*, **15**(1) (2011), 18-28.
- [2] M.C. Pol, S et al. Sensor monitoring to measure and support activities of daily living for independently living older people, *Journal of the American Geriatrics Society* (2013) 2219-2227.
- [3] L. Alpay, R. Doms, H. Bijwaard. Embedding persuasive design for self-health management systems in Dutch healthcare informatics education: Application of a theory-based method, *Health Informatics Journal* (2018), September 1-16.

Physicians' and Pharmacists' Opinions Regarding the e-Prescription Systems

Giorgos NIKOU¹ and Parisis GALLOS

*Department of Computer Science and Biomedical Informatics,
University of Thessaly, Lamia, Greece*

Abstract. The e-Prescription systems are modern and efficient prescribing tools which can affect the quality of the healthcare services. The aim of this study is to evaluate the e-Prescription systems in Greece based the users' opinions. The survey was conducted in 2019 among 157 randomly selected healthcare professionals through an on-line survey tool. The 51.2% of the sample were pharmacists and 48.8% physicians. 80.7% believe that the e-Prescription systems are reliable. More than the half of the professionals has the view that the e-Prescription systems are useful. Regarding to the ease of use, the e-Prescription systems, seem to can not fulfill sufficiently the participants' requirements.

Keywords. Electronic Prescription, Evaluation Study

1. Introduction

The electronic prescription systems could be assumed as modern and efficient prescribing tools for drugs and other medical product or services [1]. The usage of the e-Prescription systems has significantly reduced the errors during the prescription process as well as, led to a better control of the health costs [2]. The evaluation of these systems seems to be important as it can affect the system functionality and the users' satisfaction [3,4]. Today, two e-prescription systems are in use in Greece. The IDIKA's system is used for drug and diagnostic tests prescription and the EOPYY's system for all the other prescribing items. The aim of this pilot study is to record the physicians' and pharmacists' opinions regarding the above e-Prescription systems in order to evaluate the systems' reliability, usefulness and ease of use.

2. Methods

To achieve the aim of this study, a self-developed questionnaire was used. The questionnaire included questions about the participants' age, profession, years of experience, their opinions regarding the mistakes during the prescription process due to the systems, the reliability, the usefulness and the ease of use of the e-prescription systems, and the participant's cost of usage for the e-prescription services. The survey was conducted in Greece in 2019 among 157 randomly selected Physicians and

¹Corresponding Author, Giorgos Nikou, Student at the Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece; E-mail: gnikou07@gmail.com.

Pharmacists through an on-line survey tool. Descriptive statistics and correlations between the participants' profession and their opinions were examined using SPSS.

3. Results and Discussion

After the statistical analysis, the 51.2% of the sample were pharmacists and 48.8% physicians. The mean of the age was 43.46 years old. About the professional experience, the 23.3% of them have more than 10 years. As 39% of all mistakes on the prescription process are due to the e-Prescription system malfunctions, the 79.1% of the participants believe that the last two years the number of errors, due to system malfunctions, was decreased. Meanwhile, the 80.7% think that the e-Prescription systems are reliable. For IDIKA's system, the 86% of the participants assumes that the system is useful and 73.8% accepts that it is easy to use. On the other hand, for the EOPYY's system, the 56.9% of the participants believe that the system is useful and only 22.7% thinks that it is easy to use. Additionally, the mean of the annual cost of usage for the e-prescription services is 554.78 euro. Comparing the participants' profession with their opinions regarding the e-Prescription systems, all comparisons were not confirmed ($p > 0.05$), except the comparison between the profession and the perceived ease of IDIKA's system use ($p < 0.01$). Finally, the pharmacists believe that the IDIKA's prescription system is easy to use, way more than physicians. No other differences in opinions between physicians and pharmacists were found.

4. Conclusions

According to the aforementioned results, although the participants believe that a large percentage of the prescription errors are related to the e-Prescription systems, they trust the systems, and these systems are assumed reliable by the majority of the participants. The majority of the professionals have the view that the e-Prescription systems are useful. Regarding to the ease of use, the e-Prescription systems, does not fulfill sufficiently the participants' requirements and on one of the system, pharmacists seem to be more familiar with the systems usage rather than the physicians. Future work may include a wider scale survey using structured and valid measuring tools in order to record more accurately the healthcare professionals' opinions about the e-Prescription systems on international level.

References

- [1] A. Porterfield, K. Engelbert, A. Coustasse, Electronic prescribing: improving the efficiency and accuracy of prescribing in the ambulatory care setting, *Perspectives in Health Information Management* **11** (2014).
- [2] M.A. Fischer, C. Vogeli, M. Stedman, T. Ferris, et al. Effect of electronic prescribing with formulary decision support on medication use and cost, *Archives of Internal Medicine* **168**(22) (2008), 2433-2439.
- [3] W.S. Tan, J.S. Phang, L.K. Tan, Evaluating user satisfaction with an electronic prescription system in a primary care group, *Annals Academy of Medicine Singapore* **38**(6) (2009), 494.
- [4] A. Kontogiannatou, J. Liaskos, P. Gallos, J. Mantas, Usefulness, Ease of Use, Ease of Learning and Users' Satisfaction of E-Prescription and E-Appointment Systems for Primary Health Care, *Studies in Health Technology and Informatics* **262** (2019), 210-213.

PoleSat's Modelling a Reorganization of Vascular Catheterization Provision at the Territorial Hospital Grouping Level

Anne QUESNEL-BARBET ^{a,f,1}, Julien SOULA ^b, Erik-André SAULEAU ^c, Pierre PARREND ^c, Pierre BAZILE ^d, François DUFOSSEZ ^e, Gilles MAIGNANT ^f, Pascal STACCINI ^f, Frédéric ALBERT ^g and Arnaud HANSSKE ^h

^a Lille University Hospital (CHU), MRRC, Public Health, Lille, Nord, France.

^b CERIM Lab - EA 2694-Medical Faculty of Lille University, Lille, Nord, France.

^c ICube Lab – CNRS – CSDC-UNESCO-Unitwin, Strasbourg University, B-R, France.

^d Geographic Information Association (AFIGEO), Saint-Mandé, Val de Marne, France.

^e Medical Information Dept, Béthune Hospital Center, Béthune, Pas-de-Calais, France.

^f RETINES Lab - IRIS Dept - Medical Faculty of Nice University, Nice, PACA, France.

^g Care Supply General Direction - DGOS, Ministry of health, Paris, IdF, France.

^h KASHMIR-DataReuse Lab, Catholic Lille University (UCL), Lille, Nord, France.

Abstract. Since a French organization (2016) has defined “the territorial hospital groupings”, public hospitals must share medical-economic knowledge and decision-makers expect prospective analyses. PoleSat aims, quick hospital-catchment area modellings, completed by population analyses. Modellings are based on “diagnostic and interventional vascular catheterizations” acts and Nouvelle-Aquitaine, and they are carried out 3 times, through the graphical user interface’s main-setting values, coupled with 3 activity-scenarios. Scenario results cannot confirm the NA02-Atlantique’s H0. The experts have approved PoleSat’s method as a robust help-tool; therefore they project to repeat its usages.

Keywords. Models, Gravitation, Geographic Mapping, Management Information Systems, Diagnosis-Related Groups (DRG), Decision Making, Computer-Assisted

1. Introduction

Since a French organization (2016) has defined “the territorial hospital groupings THGs” i.e. a new sanitary level for health management, public hospitals must share medical-economic information [1]; also, keeping up with this movement, health decision-makers need prospective activity/recruitment analyses; thus, to help decision-makers, PoleSat tailored IT tool [2] aims quick-easy modelling of hospital catchment area (CA), completed by population analyses (per scenario). Two main hospital centers (CH) are targeted to predict the activity transfer impacts between Rochefort - and

¹ Anne Quesnel-Barbet, CHU of Lille, MRRC, 6 rue du Pr Laguesse, CS 70001, 59037 Lille cedex, France; E-mail : aquesnel@univ-lille.fr

Saintes – CH, and to accept/reject the *NA02 – Atlantique 17 THG's* recruitment hypothesis H0.

2. Materials and methods

PoleSat allows to upload its own health *ready-input file*, [2] and to ensure a general data protection regulation compliance. Other variables are added [3] to obtain localization & geographic limits. Modellings are applied on “diagnostic and interventional vascular catheterizations (DIVC)” activity from DRG-PMSI 2018 [1], and concern 2 Nouvelle-Aquitaine's THGs.

Initial, transfer, and removal activity-scenarios are considered to confirm/refute H0: “High positive activity gain of 100% is expected for Saintes, under a transfer of 617 acts from Rochefort - to Saintes”. Modellings are carried out 3 times, through the graphical user interface's main-settings (P1: 2, P2: 54, P5: 2, & P8: 15), coupled with PMSI-ready-input files tied to each activity-scenario.

3. Results

The 3 activity-scenario modellings are presented in Figure 1.

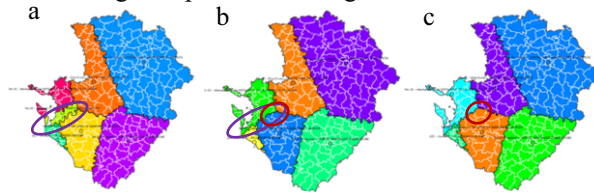


Figure 1. a- “initial scenario”, Rochefort’s CA inside a purple-ellipse appears in light-green. b- “removal scenario”, Rochefort’s CA has disappeared & almost all its CA is eaten by La Rochelle’s CA. c- “transfer scenario”, inside a red-circle Rochefort’s activity transfer to Saintes. Saintes’ CA has weakly increased in “c” in comparison with the similar area in “b”

4. Discussion & Conclusion

Each hospital activity around Rochefort increases proportionally to the covered-population growth; also, the transfer does not lead to the expected decrease of DIVC in Rochefort’s area; thus, La Rochelle could attract 2 to 3 times more patients than Saintes; therefore, the outcomes cannot confirm H0. Reduce activity leak of Saintes appears evident for the experts, which have validated Polesat (distance & mass) as a robust help tool, and project to repeat its usage to other health planning cases.

References

- [1] The law 2016-41: our health system modernization. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000031912641&categorieLien=id>
- [2] Quesnel-Barbet A, Soula J, Sauleau EA, Parrend P, Bazile P, Dufosse F, Hansske A (2019) Polesat-Web-2018: A Simulation IT Tool with Immediate Prospective and Strategic Views of Hospital Spatial Planning. Stud Health Technol Inform. <https://doi.org/10.3233/SHTI190633>
- [3] Quesnel-Barbet A (2019) Data repository used for PoleSat. <http://pr-polesat2018/polesat2019-metier/doc/referentiel.xlsx>.

Predicting Diagnosis Code from Medication List of an Electronic Medical Record Using Convolutional Neural Network

Jakir Hossain Bhuiyan MASUD^a, Ming-Chin LIN^{a,b,c,1}

^a*Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan*

^b*Division of Neurosurgery, Department of Surgery, Shuang-Ho Hospital, Taipei, Taiwan*

^c*Taipei Neuroscience Institute, Taipei Medical University, Taipei, Taiwan*

Abstract. Automated coding and classification systems play a role in healthcare for quality of care. Our objective was to predict diagnosis code from medication list of electronic medical record (EMR) using convolutional neural network (CNN). We collected the clinical note from outpatient department (OPD) of Wanfang hospital, Taiwan of 2016 and used three physicians from three departments. The dataset was split into two parts, 90% for training and 10% for test cases. We used medication list as input and International Statistical Classification of Diseases 10 (ICD 10) code as output. After data preprocess, we used word2vector CNN to predict ICD 10 code. This study shows all the three physicians from three departments achieved better performance. The best performance of model was a physician from cardiology department achieved precision 69%, recall 89% and F measure 78%. We need to include more component such as text data, lab report for evaluation.

Keywords. Clinical note, International Statistical Classification of Diseases 10 (ICD 10), medication, convolutional neural network

1. Introduction

Electronic medical record (EMR) is routinely collected heterogenous data. Each record should have complete information of problem list to establish diagnosis and abnormal findings [1]. However, it is hard to create accurate predictive model, because of heterogenous data, data quality and label availability [2]. Medical coder put International Statistical Classification of Diseases 10 (ICD 10) code manually in the EMR and it's labor intensive and time consuming. A study used 17 conditions for identifying problem lists of patients based on medication and laboratory results [3]. Natural language processing (NLP) uses for different problem transformation approaches and employing suitable label calibration and ranking methods [4]. Pre-trained word vectors use for sentence-level classification on CNN [5]. Our aim was to explore whether we can automatically assist physician to predict diagnosis code from medication list of EMR using CNN.

¹ Corresponding author: E-mail: arbiterski@gmail.com

2. Method

Dataset: This study is a retrospective cohort of patients. We collected clinical notes of outpatient department (OPD) from the EMR of Wanfang Hospital, Taiwan of 2016, consists of SOAP, medication, diagnosis lists. The total medical records were 10,364 where 9,583 (90%) for training and 781 (10%) for test set that include three Physicians.

Experimental setup: We consider the task as multi-label classification problem, where the input is medication list and the output is the ICD-10 label of clinical note. We implemented word embedding word2vector (w2v) CNN for multi-label classification task using python, Keras. We used Sigmoid activation function at the end of CNN. We used ranking method for getting top 10, top 5 diagnosis using the dataset. This ranking can identify the best diagnosis chronologically. Initially it takes the tokenization, use w2v and then use CNN. CNN used with Keras and weight parameters are optimized using Adam method that is a variant of the stochastic gradient descent (sgd) algorithm.

Evaluation: The evaluation was done by precision, recall and F measure.

3. Result

Our study shows the best performance of CNN model for a physician from cardiology department as precision, recall and F measure were 69%, 89% and 78% respectively. The predicted ICD 10 using probability that identify the ranking where E782 Mixed hyperlipidemia is the top (probability 0.9972247) of a diagnosis list.

4. Conclusion

We find better model performance as well as better ICD 10 prediction chronologically through probability. However, we need to explore the evaluation by all physicians and include text data, lab report in future tasks.

References

- [1] Burton, M. M., Simonaitis, L., & Schadow, G. (2008). Medication and indication linkage: a practical therapy for the problem list?. In *AMIA Annual Symposium Proceedings* (Vol. 2008, p. 86). American Medical Informatics Association.
- [2] Xiao, C., Choi, E., & Sun, J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, 25(10), 1419-1428.
- [3] Wright A, Pang J, Feblowitz JC, et al. A method and knowledge base for automated inference of patient problems from structured data in an electronic medical record. *Journal of the American Medical Informatics Association* : JAMIA 2011;18(6):859-67 doi: 10.1136/amiajnl-2011-000121[published Online First: Epub Date]].
- [4] Kavuluru R, Rios A, Lu Y. An empirical evaluation of supervised learning approaches in assigning diagnosis codes to electronic medical records. *Artificial intelligence in medicine* 2015;65(2):155-66 doi: 10.1016/j.artmed.2015.04.007[published Online First: Epub Date].
- [5] Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; 2014. Association for Computational Linguistics.

Prediction of ROSC After Cardiac Arrest Using Machine Learning

Nan LIU^{a,b,1}, Andrew Fu Wah HO^{b,c}, Pin Pin PEK^{b,c}, Tsung-Chien LU^d, Pairoj KHRUEKARNCHANA^e, Kyoung Jun SONG^f, Hideharu TANAKA^g, Ghulam Yasin

NAROO^h, Han Nee GANⁱ, Zhi Xiong KOH^c, Huei-Ming MA^d, Marcus ONG^{b,c}

^a Health Services Research Centre, Singapore Health Services, Singapore

^b Health Services and Systems Research, Duke-NUS Medical School, Singapore

^c Department of Emergency Medicine, Singapore General Hospital, Singapore

^d Department of Emergency Medicine, National Taiwan University Hospital, Taiwan

^e Department of Emergency Medicine, Rajavithi Hospital, Thailand

^f Seoul National University College of Medicine, Republic of Korea

^g Graduate School of EMS System, Kokushikan University, Japan

^h ED-Trauma Centre, Rashid Hospital, Dubai, United Arab Emirates

ⁱ Accident & Emergency Department, Changi General Hospital, Singapore

Abstract. Out-of-hospital cardiac arrest (OHCA) is an important public health problem, with very low survival rate. In treating OHCA patients, the return of spontaneous circulation (ROSC) represents the success of early resuscitation efforts. In this study, we developed a machine learning model to predict ROSC and compared it with the ROSC after cardiac arrest (RACA) score. Results demonstrated the usefulness of machine learning in deriving predictive models.

Keywords. Out-of-hospital cardiac arrest, ROSC, machine learning, random forest

1. Introduction

The return of spontaneous circulation (ROSC) serves as an indicator of the success of early resuscitation in out-of-hospital cardiac arrest (OHCA) patients. Given the global burden of OHCA and disparity among healthcare systems, a simple tool is needed to enable comparison of different EMS systems worldwide. The ROSC after cardiac arrest (RACA) score [1] developed in Germany is one solution, which uses variables that are available after EMS arrival to assess the probability of ROSC through logistic regression. In this study, we aimed to create a machine learning model to predict ROSC.

2. Method

We used the data collected from Pan-Asian Resuscitation Outcomes Study (PAROS) [2]. The data consists of OHCA victims recruited between January 2009 and December 2012 from Japan, South Korea, Malaysia, Singapore, Taiwan, Thailand, and United Arab Emirates (UAE). In this study, we excluded paediatric cases (<18 years old), non-EMS transported cases, and those with a “do not resuscitate” order. We used random forest

and following variables to build a prediction model: gender, age, cardiac arrest aetiology, witness, location, initial rhythm, bystander cardiopulmonary resuscitation (CPR), and EMS arrival time. We randomly selected 70% data for random forest model training and the remaining 30% data for testing the model and the RACA score.

3. Results

A total of 63,059 patients were used in our analysis, of which 5,190 (8.2%) patients had ROSC at scene or en route to ED. Figure 1 shows receiver operating characteristic (ROC) curves generated by the random forest model and the RACA score. Area under the ROC curves were 0.76 and 0.73 for random forest and RACA score, respectively.

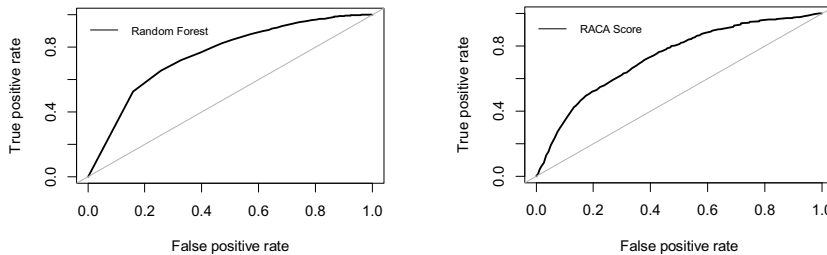


Figure 1. ROC curves of the random forest model and the RACA score.

4. Discussion

Our study showed that a machine learning model (random forest) outperformed RACA score in predicting ROSC among OHCA patients. The results demonstrated the power of machine learning in deriving predictive tools. Overall, the discriminatory abilities of both machine learning model and RACA score (0.76 and 0.73 AUC) were comparable to the original RACA score (0.73 AUC) on German registry patients. Our study proved the applicability of the RACA score in Pan-Asian countries and illustrated the possibility of applying machine learning algorithms in developing new predictive tools.

5. Conclusion

Machine learning is worthy of further investigation in developing powerful predictive tools for OHCA and other medical applications.

References

- [1] J.T. Grasner, P. Meybohm, R. Lefering, J. Wnent, J. Bahr, M. Messelken, T. Jantzen, R. Franz, J. Scholz, A. Schleppers, B.W. Bottiger, B. Bein, and M. Fischer, ROSC after cardiac arrest—the RACA score to predict outcome after out-of-hospital cardiac arrest, *Eur Heart J* **32** (2011), 1649-1656.
- [2] M.E. Ong, S.D. Shin, H. Tanaka, M.H. Ma, P. Khruengkarnchana, N. Hisamuddin, R. Atilla, P. Middleton, K. Kajino, B.S. Leong, and M.N. Khan, Pan-Asian Resuscitation Outcomes Study (PAROS): rationale, methodology, and implementation, *Acad Emerg Med* **18** (2011), 890-897.

Predictors of Hospital Admissions Based on Routinely Collected Everyday Observations in a Community Care Setting

Mads Nibe STAUSHOLM^{a,1}, Pernille Heyckendorff SECHER^a and Ole Kristian HEJLESEN^a

^a*Department of Health Science and Technology, Aalborg University, Denmark*

Abstract. The aim of this study was to identify predictors for hospital admissions in community-dwelling adults based on routinely collected community data. Univariate logistic regression analyses were performed to assess each variable's ability to predict preventable and all cause hospital admissions.

Keywords. Prediction, primary health care, elderly, home care, hospital admission

1. Introduction

Population ageing challenges the existing healthcare systems as the older population are at higher risk of chronic diseases that require long-term care (1), and more frequent hospital admissions (2). Consequently, avoiding unnecessary hospital admissions have become a priority, as both the primary- and secondary healthcare sector cannot keep up with the rising demand (3). The Triage Changing Table has been recommended by The Danish Health Authority for early detection of acute events and prevention of preventable hospital admissions based on home care providers' everyday observations of clients' health states (4,5). No studies have examined the association between data obtained from the Triage Changing Table Database and hospital admissions. The aim of the present study was to identify predictors for preventable and all cause hospital admissions in community-dwelling adults based on everyday health state observations.

2. Methods

First paragraph. Data was retrospectively collected from The Patient Administrative System used in the North Denmark Region, which includes information about hospital admissions, and The Triaged Changing Table Database. 4,410 subjects aged 65 or above residing in Aalborg Municipality, Denmark, were included in the study. The outcomes were preventable and all cause, acute hospital admissions defined as an unplanned overnight stay in hospital. Prior to each acute hospital admission, we defined a prevention, event, and control period with length of three, 14, and 14 days, respectively. A total of 279 and 261 features were constructed from the Triage Changing Table variables for the

¹ Corresponding Author.; E-mail: mns@hst.aau.dk.

preventable and all cause hospital admission analyses, respectively. Univariate logistic regression analyses were performed to assess each variable's ability to predict the outcome. Five-fold cross-validation was applied to reduce the risk of overfitting and increase generalizability to new data.

3. Results

A total of 68 control and event period pairs from 53 subjects was included in the all cause hospital admission analysis. 19 control and event period pairs from 18 subjects were included in the preventable hospital admission analysis. The most discriminative predictor variable for all cause hospital admissions was the difference in total sum of item codes between the first and last registration in each period (OR = 1.20; 95% CI = 0.94, 1.55; $p = .14$; $AUC \pm STD = .60 \pm .12$). For the preventable hospital admissions, the most discriminative predictor variable was the number of Triage Changing Table registrations (OR = 1.05; 95% CI = 0.92, 1.19; $p = .48$; $AUC \pm STD = .65 \pm .16$). No predictors showed statistically significant associations with the outcome.

4. Discussion

This study used only routinely collected data extracted from administrative databases in a community care setting, which compared to self-reported data tends to have greater predictive performance (6). The reported AUCs indicate some discriminative potential, however, no variables had statistically significant association to the outcome, which may be due to the relatively low sample sizes. The current data segmentation strategy limited the number of subjects and control/event periods. A less restrictive segmentation approach will likely increase the number of event and control periods.

In conclusion this study has demonstrated that data collected using the Triage Changing Table may be useful in predicting impending hospital admissions. However, further research is needed to evaluate the true predictive potential of this tool.

References

- [1] He, W, Goodkind D, Kowal P. *An Aging World: 2015*, International Population Reports Aging (Albany NY). 2016;(March)
- [2] Vest-Hansen B, Riis AH, Sørensen HT, Christiansen CF. Acute admissions to medical departments in Denmark: Diagnoses and patient characteristics. *Eur J Intern Med* [Internet]. 2014; 25(7):639–45.
- [3] Mitzner TL, Beer JM, McBride SE, Rogers WA, Fisk AD. Older Adults' Needs for Home Health Care and the Potential for Human Factors Interventions. *Proc Hum Factors Ergon Soc Annu Meet*. 2009;53(11):718–22.
- [4] The Danish Health Authority. Tidlig opsporing af sygdomstegn hos ældre medicinske patienter: Værktøjer til hverdagsobservationer. 2013;1–24.
- [5] Stausholm MN, Secher PH, Cichosz SL, Hejlesen OK. Predicting Preventable Hospitalizations among Elderly Recipients of Home Care: A Study Protocol. In: *Proceedings from The 16th Scandinavian Conference on Health Informatics 2018*, Aalborg, Denmark August 28–29, 2018. 2018. p. 75–9.
- [6] Wallace E, Stuart E, Vaughan N, Bennett K, Fahey T, Smith SM. Risk Prediction models to predict emergency hospital admission in community-dwelling adults: A systematic review. *Med Care*. 2014;52(8):751–65.

Preliminary Evaluation of a mHealth Coaching Conversational Artificial Intelligence for the Self-Care Management of People with Sickle-Cell Disease

David-Z. ISSOM^{a,1}, Jessica ROCHAT^a, Gunnar HARTVIGSEN^b, Christian LOVIS^a

^a Division of Medical Information Sciences, Geneva University Hospitals, Switzerland

^b UiT The Arctic University of Norway, Tromsø, Norway.

Abstract. Adherence to the complex set of recommended self-care practices among people with Sickle-Cell Disease (SCD) positively impacts health outcomes. However, few patients possess the required skills (i.e. disease-specific knowledge, adequate levels of self-efficacy). Consequently, adherence rates remain low and only 1% of patients are empowered enough to master the self-care practices. Health coaching and therapeutic patient education have emerged as new approaches to enhance patients' self-management and support health behavior changes. This preliminary feasibility study examined patients' perceived usefulness of the information provided by a chatbot we developed following patient-important requirements collected during our preliminary studies. Participants tested the chatbot and completed a post-test survey. A total of 19 patients were enrolled and 2 withdrew. 15 respondents (15/17, 88%) gave a score of at least 3/4 to the question "The chatbot contains all the information I need". Results suggest that mHealth coaching apps could be used to promote the knowledge acquisition of recommended health behaviors related to the prevention of SCD main symptoms.

Keywords. Sickle-Cell Disease, Chatbots, mHealth, Patient Empowerment.

1. Introduction

Sickle-Cell Disease (SCD) is the most prevalent monogenic disorder worldwide [1,2]. The major clinical manifestations are hemolytic anemia and vaso-occlusive pain crises (VOC) [3]. Triggers of VOCs are numerous and include inadequate eating behaviors, stress, infections, dehydration, fatigue [4]. To prevent them, patients require high cognitive capabilities, good disease-specific knowledge and high self-efficacy levels [5]. Yet, literature suggests that only 1% of SCD patients are empowered enough to fully self-manage [6]. Mobile health (mHealth) apps, because of their relatively low cost and scalability, can offer a potential route to support SCD self-management. Studies have shown that Conversational Artificial Intelligence (AI) such as chatbots to encourage people with Diabetes' behavior changes, had a high acceptability and high user engagement [7]. To our knowledge, no work has been done to design such tools for the specific needs of people with SCD. This preliminary study aims at evaluating patients' perceived usefulness of the information provided by a fully automated coaching app designed to empower patients through the learning of recommended self-care practices.

¹Corresponding Author; E-mail: david.issom@unige.ch.

2. Methods

To be part of the study, applicants had to be diagnosed with SCD. Patients had to be at least 16 years old and be able to understand French. People who had been cured (e.g., bone marrow transplantation) were excluded. Finally participants had to possess a smartphone. Patients have been recruited through patient associations in Guadeloupe during the World Sickle-Cell Awareness Day. All participants gave informed consent before the evaluation and all responses were anonymous. The instrument has been developed following patient-important requirements collected during our previous work [8,9]. Participants tested the chatbot and completed a post-test, the 9-items usefulness scale for patient information material (USE) survey.

3. Results

19 patients participated to the test (9 females and 8 males). 17 completed the evaluation. 2 females withdrew because of problems with their smartphones. Patients were aged between 19 and 59 years old. 15 respondents (15/17, 88%) gave a score of at least 3/4 to the question “The chatbot contains all the information I need” for a total score of 54/68. 10 participants (10/17, 58%) gave a score of at least 3/4 to the question “The chatbot encouraged me to be more active in order to improve my condition”, for a total of 51/68.

4. Discussion

Results suggest that chatbots for health coaching can be used to promote the knowledge acquisition of recommended self-care practices related to the prevention of VOCs.

References

- [1] UN GENERAL ASSEMBLY ADOPTING CONSENSUS ON SCD | Meetings Coverage and Press Releases [Internet]. [cited 2019 Feb 26]. Available from: <http://www.webcitation.org/76VKrj7nL>.
- [2] Sedrak A, Kondamudi NP. Sickle Cell Disease. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2019 [cited 2020 Jan 6]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK482384/>.
- [3] Buchanan G, Vichinsky E, Krishnamurti L, et al. Severe SCD—pathophysiology and therapy. *Journal of the American Society for Blood and Marrow Transplantation*. 2010;16:S64–7.
- [4] Khan SASA, Damanhour G, Ali A, et al. Precipitating factors and targeted therapies in combating the perils of sickle cell disease— A special nutritional consideration. *Nutrition & metabolism*. 2016;13:50.
- [5] Jenerette CM, Lauderdale G. Successful Aging with Sickle Cell Disease: Using Qualitative Methods to Inform Theory. *Journal of theory construction & testing*. 2008;12:16–24.
- [6] Ballas SK. Self-management of SCD: a new frontier. *Journal of the National Medical Association*. 2010;102:1042–3.
- [7] Hansel B, Giral P, Gambotti L, et al. A Fully Automated Web-Based Program Improves Lifestyle Habits and HbA1c in Patients With Type 2 Diabetes and Abdominal Obesity: Randomized Trial of Patient E-Coaching Nutritional Support (The ANODE Study). *Journal of medical Internet research*. 2017;19:e360.
- [8] Woldaregay AZ, Issom D-Z, Henriksen A, et al. Motivational Factors for User Engagement with mHealth Apps. *Studies in health technology and informatics*. 2018;249:151–157.
- [9] Issom D-Z, Zosso A, Ehrler F, et al. Exploring the Challenges and Opportunities of eHealth Tools for Patients with Sickle Cell Disease. *Stud Health Technol Inform*. 2015;216:898.
- [10] Hölzel LP, Ries Z, Dirmaier J, et al. Usefulness scale for patient information material (USE) - development and psychometric properties. *BMC Med Inform Decis Mak*. 2015 [cited 2019 Oct 15];15.

Preliminary Evaluation of Market Mechanism-Based Bed Allocation System

Shusuke HIRAGI^{1a,b}, Osamu SUGIYAMA^a, Jun HATANAKA^b, Shosuke OHTERA^{a,b},
Goshiro YAMAMOTO^a, Kazuya OKAMOTO^{a,b}, Masayuki NAMBU^a, and
Tomohiro KURODA^{a,b}

^a *Kyoto University Hospital, Kyoto, Japan*

^b *Department of Social Informatics, Graduate School of Informatics, Kyoto University,
Kyoto, Japan*

Abstract. Effective bed management is important for hospital management. Until now, bed allocation process is generally controlled by administrative staffs in centralized manner but it is not always effective. In the present study, we proposed and evaluated new method for bed allocation applying market mechanism via token. Evaluation was performed with newly-developed game-type simulation. Nurse managers as research participants played it and answered for survey. The result showed that the proposed method can be useful with appropriate operational design.

Keywords. Hospital Management, Bed Occupancy, Resource Allocation

1. Introduction

Effective bed allocation is important for hospital management [1], as bed capacity is central part of hospital's competency. Current bed allocations are generally executed by administrators with centralized manner, but such management tends to result in front-line staffs' decreased job satisfaction and even invokes their resignation [2]. Considering that healthcare is labor-intensive industry, current state can be seen as ineffective and we need improvement. Conventional discussion in economics has argued that centralized decision making such as collectivist society cannot be efficient comparing to decentralized, market mechanism-based society. We here propose a novel bed allocation system based on market mechanism in which physicians who needs beds "select" and "buy" "price-tagged" impatient beds from nurse managers through transactions with token. We implemented preliminary game-like computer simulation and performed evaluation of the system through survey to nurse managers.

2. Method

We regarded bed allocation process as transactions of nursing services between physicians and nursing staffs. In particular, wards offer prices of their nursing services for physicians planning patient admission, and physicians buy them considering prices and their preferences. By doing so, we expected improved bed allocation from the

¹ Corresponding Author, E-mail: hiragi.shusuke.4x@kyoto-u.ac.jp

viewpoint of hospital management. To evaluate the system, we developed web-based simulation game in which three players were required to act as nurse managers of a hospital using aforementioned bed allocation system. Players were asked to set prices for patients as nurse managers followed by computer agent acting as physician selected the cheapest ward. The game was set to end when the tenth cycle ended. Next, we recruited three nurse managers from a single university affiliated hospital in Japan. They were asked to play simulation game described above, assuming that they are carrying out their responsibilities as managers in the in-game world. As evaluation, we recorded price fluctuation and bed allocation results. In addition, we asked participants to answer several questionnaires.

3. Results

Research participants were nurse managers in active duty and all of them were female. Their work experiences were 20-34 years as nurse and 5-14 years as nurse manager. The prices of beds fluctuated according to wards bed occupancy. Answers of questionnaires were summarized as follows; participant A was generally skeptical to the system, while participant C was favorable. Even though their answers were diverse, they evaluated “patient-centeredness” of the system negatively while “fast decision making” neutrally or positively. When they were asked about preferred reward for collected tokens, they answered as follows: bonus, additional human resource, job training allowance, welfare expenses.

4. Discussion

In the present study, we proposed a new method for bed allocation and evaluated it with game-like simulation system. Simulation system showed slight association between bed occupancy and price for diseases, indicating nurse managers’ strategies considering surrounding situations. Referring conventional theories in economics, such rationalities of each stakeholders are expected to result in overall optimization [3]. In addition, participants showed their concern that the system could not be good for patient care because it does not reflect caregivers’ specialty. In addition, appropriate rewards for the tokens is said to be the key factor for the effective implementation for token economy [4]. Our survey also showed two of three participants preferred additional human resources. This result indicated that hospital administrators will be able to manage the system via controlling human resource allocations. In conclusion, token-based bed allocation system based on market mechanism approach can be effective when it is used with appropriate operational design.

References

- [1] Forneas N. Improving hospital bed management with AI. Vol 2019: IBM; 2018.
- [2] Yamaguchi K, Ikeda S, Muto M, Narita T. Using Data Mining to Investigate the Factors Influencing Nurses to Quit Their Jobs—From the Perspectives of Bed Management, Working Environment and Job Satisfaction—. *Journal of Japanese Association for Health Care Administrators*. 2013;7:55-65.
- [3] Davies, W. (2014). Neoliberalism: A Bibliographic Review. *Theory, Culture & Society*, 31(7–8), 309–317. <https://doi.org/10.1177/0263276414546383>
- [4] Ivy JW, Meindl JN, Overley E, Robson KM. Token Economy: A Systematic Review of Procedural Descriptions. *Behavior modification*. 2017;41:708-737.

Preliminary Results from a Pragmatic Clinical Trial of MyPEEPS Mobile to Improve HIV Prevention Behaviors in Young Men

Rebecca SCHNALL^{a,1} Lisa KUHNS^b Cynthia PEARSON^c Josh BRUCE^d D. Scott BATEY^e Asa RADIX^f Uri BELKIND^f Marco A HIDALGO^g Sabina HIRSHFIELD^h Sarah GANZHORN^a Robert GAROFALO^b

^a Columbia University School of Nursing, NY, NY USA; ^b Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA; ^c 3 University of Washington, Seattle, WA, USA; ^d Birmingham AIDS Outreach, Birmingham, AL, USA; ^e University of Alabama Birmingham, AL, USA; ^f Callen Lorde Community Health Center, NY, NY, USA; ^g Children's Hospital of Los Angeles, Los Angeles, CA, USA; ^h SUNY Downstate Medical Center, Brooklyn, NY, USA

Abstract. Our study team developed the MyPEEPS Mobile App for improving HIV prevention behaviors in diverse young men. We conducted a randomized controlled trial and evaluated the preliminary outcomes in the first half (N=350) of our intended study sample. Higher self-efficacy for HIV prevention behaviors ($p=0.0042$) and more recent HIV tests in the past 3 months ($p=0.0156$) were reported by the intervention group compared to control. Numbers of condomless anal sex acts were lower among the intervention group for both insertive anal sex acts ($p=0.0283$) and receptive anal sex acts ($p=0.0001$). Preliminary results indicate that some sexual risk behaviors were reduced among the intervention group in the preliminary analytic sample.

Keywords. mHealth, digital personalized health, young men, HIV prevention

1. Introduction

Increased HIV incidence continues among young men who have sex with men (YMSM), with rates especially disproportionate among racial and ethnic minority YMSM in the US. Yet, there remains a dearth of evidence-based HIV prevention interventions for YMSM – and few that address racially- and ethnically-diverse YMSM. We employed a community-engaged research approach to design the MyPEEPS Mobile App targeting ethnically-diverse YMSM. MyPEEPS delivers information through 21 activities comprised of: didactic content, graphical reports, videos, and true/false and multiple-choice quizzes. [1] This ongoing randomized controlled trial is testing the efficacy of MyPEEPS, an evidence-based intervention for

¹ Corresponding Author: E-mail: rb897@columbia.edu.

reducing sexual risk behaviors (e.g., sex without condoms) and strengthen protective factors (e.g., increasing self-efficacy around safer sex) among racially and ethnically diverse YMSM across the US.

2. Methods

Young males (ages 13-18) who report same-sex attraction and have kissed another male or plan to have sexual activity with another male in the next year were recruited from across the US (Figure 1). Enrolled participants provide demographic, psychosocial, and sexual behavior data at baseline. The intervention group receives the MyPEEPS Mobile App at the baseline visit, with follow-up assessments at 3-month, 6-month, and 9-month time points. The control (delayed intervention) group completes 3-month, 6-month, 9-month, and 12-month assessments. Regression analyses assessed data collected from a sub-sample of 350 baseline visits and 271 matched 3-month visits.



Figure 1. MyPEEPS Mobile social media recruitment banner.

3. Results

Higher self-efficacy for HIV prevention behaviors ($p=0.0042$) and more HIV testing in the past 3 months ($p=0.0156$) were reported by the intervention group compared to control. Numbers of condomless anal sex acts were lower among the intervention group than control for both insertive anal sex acts ($p=0.0283$) and receptive anal sex acts ($p=0.0001$). The intervention group reported fewer condomless insertive anal sex acts ($p=0.0061$) but more condomless receptive anal sex acts than the control group ($p=0.0001$) while under the influence of alcohol and drugs. There were no group differences in the other HIV prevention measures.

4. Conclusions

Preliminary results indicate that some sexual risk behaviors were reduced among the intervention group in the preliminary analytic sample. Further analysis after study completion will provide more conclusive results.

References

- [1] Schnall R, Kuhns LM, Hidalgo MA, Powell D, Thai J, Hirshfield S, et al. Adaptation of a Group-Based HIV RISK Reduction Intervention to a Mobile App for Young Sexual Minority Men. *AIDS Education and Prevention*. 2018;30(6):449-62.

Preliminary Safety Analysis of a Wearable Clinic for the Early Detection of Psychotic Relapse

Ibrahim HABLI^{a,1}, Charlotte STOCKTON-POWDRELL^b, Matthew MACHIN^b, Paolo FRACCARO^b, Shon LEWIS^b and Niels PEEK^b

^aUniversity of York, United Kingdom

^bUniversity of Manchester, United Kingdom

Abstract. We discuss the preliminary safety analysis of a smartphone-based intervention for early detection of psychotic relapse. We briefly describe how we identified patient safety hazards associated with the system and how measures were defined to mitigate these hazards.

Keywords. Safety, wearables, smartphones apps, serious mental illness

1. Introduction

We focus on a smartphone-based intervention, called the Wearable Clinic, which assists in remote monitoring of patients with chronic conditions, including serious mental illness. A key function of the Wearable Clinic is to detect relapse early by identifying sudden changes in activity and behaviour, particularly for patients with schizophrenia. Relapses often result in unscheduled hospital admissions, with substantial suffering of affected individuals and their families as well as high costs for mental health services. Despite the potential benefits of the Wearable Clinic, it is a complex digital intervention in a complex clinical and social setting. This increases the risk of new unintended hazardous events that can compromise patient safety [1]. The aim of this paper is to show how the consideration of these safety concerns was incorporated into the early design of the Wearable Clinic by proactively conducting a hazard analysis in order to generate safety requirements for mitigation of the identified safety hazards.

2. Methods

We modelled the intended use of the Wearable Clinic for an early detection of psychotic relapse in an explicit use case, clearly representing the flow of activities, decisions and data. This activity has been conducted with a multidisciplinary team comprising data scientists, engineers, economists, clinicians and public contributors. This use case provided a basis for scoping and conducting a hazard and risk analysis, which is a mandatory requirement for safety standards. The safety analysis of the

¹ Corresponding Author, Ibrahim.habli@york.ac.uk.

Wearable Clinic, as scoped by our use case, was conducted using the Software Hazard Analysis and Resolution in Design (SHARD) technique [2]. SHARD is used in the safety-critical domain to assess the suitability of a proposed design of a data-intensive system and derive safety requirements for the detailed development of the design. SHARD is structured around data flows, considering inputs to the system, e.g. from sensors, and outputs, e.g. to alerting devices. SHARD uses a set of guide words (*omission*, *commission*, *early*, *late* and *value*) for identifying potential deviations from the intended behaviour of each data flow, prompting the analysts to determine plausible causes, hazardous effects and the safety requirements.

3. Results

SHARD was applied to all the data flows in the use case. Take the data flow between the decision ‘*Is risk of relapse high?*’ and the activity ‘*Inform care team*’, considering three key deviations (*omission*, *commission* and *late*). Here, the omission and late reporting of high risks of relapse represent two potential hazardous failures that have to be mitigated. For example, a potential omission cause is a common smartphone notification option that can centrally disable all notifications. This could limit the ability of the Wearable Clinic to collect user data and thus to proactively predict relapse. A mitigation measure comprises user training and greater control over central OS functions. These form explicit safety requirements for the subsequent detailed design, and the potential deployment, of the app.

4. Discussion and Conclusions

An important feature of the Wearable Clinic is that its functions are interweaved into the care pathway. With this benefit comes the challenge of identifying the safety considerations that have to be mitigated specifically by the Wearable Clinic designers. For example, is the scope of the system limited to (1) alerting the care team, (2) automatically initiating an intervention or/and (3) guiding the intervention? The greater the scope, the more safety-critical the system becomes, with more stringent regulatory constraints. Further, the confidence with which the safety requirements have to be satisfied should be proportionate to the risks posed by the technology. However, neither the safety standards nor clinical guidelines state what would be deemed as acceptable risk targets. Without these targets, it is a significant challenge for the engineers to make transparent decisions concerning the reliability of the different system components, e.g. choice of accelerometers or mobile phone platforms, in the absence of any qualitative or quantitative notion of risk acceptance.

References

- [1] I. Habli, et al, What is the safety case for health IT? A study of assurance practices in England, *Safety Science* 110 (2018): 324-335.
- [2] D. Pumfrey, *The principled design of computer system safety analyses*. Diss. Uni of York, 1999.

Primary Care Prostate Cancer Case Ascertainment

Agnieszka LEMANSKA^{a,b,1}, Sara FAITHFULL^b, Harshana LIYANAGE^d, Sophie OTTER^c, Marina ROMANCHIKOVA^a, Julian SHERLOCK^d, Nadia A. S. SMITH^a, Spencer A. THOMAS^a and Simon de LUSIGNAN^d

^aData Science Department, National Physical Laboratory, Teddington, UK

^bSchool of Health Sciences, University of Surrey, Guildford, UK

^cRoyal Surrey NHS Foundation Trust, Guildford, UK

^dNuffield Department of Primary Care Health Sciences, University of Oxford, UK

Abstract: Although routine healthcare data are not collected for research, they are increasingly used in epidemiology and are key real-world evidence for improving healthcare. This study presents a method to identify prostate cancer cases from a large English primary care database. 19,619 (1.3%) men had a code for prostate cancer diagnosis. Codes for medium and high Gleason grading enabled identification of additional 94 (0.5%) cases. Many studies do not report codes used to identify patients, and if published, the lists of codes differ from study to study. This can lead to poor research reproducibility and hinder validation. This work demonstrates that carefully developed comprehensive lists of clinical codes can be used to identify prostate cancer; and that approaches that do not solely rely on clinical codes such as ontologies or data linkage should also be considered.

Keywords. Prostate cancer, primary care, case identification, real-world evidence

1. Introduction

Prostate cancer is a chronic condition, with good prognosis but long-term consequences. Health services worldwide are stretched so it is important to understand the needs of the growing population of people affected by cancer. The increasing use of electronic health records (EHRs) can facilitate epidemiological research. However, the usability of data is limited by the lack of standards and variability in the quality of recording of cancer diagnosis and treatment. In this study, we investigate the usability of clinical codes to identify and describe a cohort of prostate cancer patients from an English database.

2. Methods

A primary care-based, retrospective cohort study using a nationally representative, English database from the Royal College of General Practitioners (RCGP), Research and Surveillance Centre (RSC) was used [1]. A December 2016 data extract was analyzed comprising 164 practices and 1,529,922 EHRs for male patients.

¹Corresponding Author, E-mail: agnieszka.lemanska@npl.co.uk

Prostate cancer diagnosis was extracted using commonly reported diagnostic codes. This included B46 (*malignant neoplasm of prostate*), B834 (*carcinoma in situ of prostate*) and HNG0200 (*cancer of the prostate*). We also included 4M01 and Xalmc (*Gleason prostate grade 5-7 (medium)*), and 4M02 and Xalmd (*Gleason prostate grade 8-10 (high)*). Date and age at diagnosis was based on the first occurrence of the clinical code. Ethnicity was extracted in five categories using an ontology-based algorithm [2].

3. Results

The average age at diagnosis was 70.8 (SD = 9.5) and the majority of men (11,586, 94.1%) were white (Table 1). Men identified with prostate cancer were spread evenly across the country (Figure 1). The most commonly extracted treatment was hormone therapy, 9324 (47.5%) men received it. The mean age at death was 80.6 (SD = 9.1).

Diagnosis was established for 19,619 (1.3%) men. The three codes: B46, B834 and HNG0200 were used for 18,356 (93.6%), 1,549 (7.9%) and 666 (3.4%) men respectively. Medium Gleason grading was recorded in 1535 EHRs and high in 562 EHRs. This enabled identification of 94 (0.5%) additional cases.

Characteristic	Mean (SD)	n (%)
Age at diagnosis (years)	70.8 (9.5)	
Ethnicity		
White		11586 (59.1)
Asian		195 (1.0)
Black		432 (2.2)
Mixed		67 (0.3)
Other		37 (0.2)
Missing		7302 (37.2)
Treatment		
Radiotherapy		634 (3.2)
Hormone therapy		9324 (47.5)
RT and HT		1970 (10.0)
None recorded		7691 (39.2)
Age at death (years)	80.6 (9.1)	
Not recorded		18170 (92.6)

Table 1. Demographic and treatment characteristics



Figure 1. Location of study participants.

4. Discussion and conclusion

Primary care-based epidemiological studies often identify cancer cases using diagnostic codes. In clinical practice, we found variations in the recording of cancer diagnosis. This work contributes to establishing standards for primary care cancer case ascertainment.

References

[1] A. Correa, et al., Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile, *BMJ Open* **6** (2016).
[2] Z. Tippu, et al., Ethnicity Recording in Primary Care Computerised Medical Record Systems: An Ontological Approach, *J Innov Health Inform* **23** (2017), 920.

Proposal of an Architecture for Terminology Management in a Research Project

Sara MORA^{a,1}, Sumit MADAN^b, Stephan GEBEL^b, Mauro GIACOMINI^a

^a*Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Italy*

^b*Fraunhofer Institute for Algorithms and Scientific Computing SCAI Department of Bioinformatics Schloss Birlinghoven, Sankt Augustin, Germany*

Abstract. Clinical and medical knowledge evolve and this causes changes in concepts and terms that describe them. The objective of this work is to formally present an ontology-based standard architecture that will be used in the scenario of neurodegeneration research to maintain terminologies and their relations updated and coherent over the time. The proposed structure is composed by three elements that will allow the user to do a list of operations on the terminology resources explicitly contemplated by the *Common Terminology Service Release 2* (CTS2).

Keywords. CTS2, ontologies harmonization.

1. Introduction

As in all scientific domains, clinical and medical knowledge evolve over time. In a scenario of multiscale and multimodal data integration it is mandatory to use shared semantics because of the need to deal with heterogeneous datasets [1]. In order to correctly manage this type of semantic framework, it is important to consider that the medical knowledge evolution causes the change of concepts and of the terms used to define them. For this reason, it is necessary to manage historical evolution and versioning [2]. So, in the scenario of collaboration projects, it is essential to use both controlled vocabularies and standard infrastructures to guarantee both high-quality data collection and interoperability between partners.

2. Methods

As defined in [3], ontologies have been originally developed from the need to represent and to communicate a domain's knowledge unambiguously. As the project is linked to neurodegeneration research, a group of specific ontologies have been selected. This will not only guarantee the correct dialogue between homogeneous datasets, but it will also support the information extraction workflows for “cause-and-effect”-type relationships. Ontologies are not isolated; indeed, it is possible to define links between them in a process of ontologies harmonization. The four ways to reach this goal are: matching, mapping, alignment and merging. In the context of this work, it will be

¹ Corresponding Author, E-mail: sara.mora@edu.unige.it

necessary to guarantee the correct mapping between the ontologies. As detailed in [4], the objective of CTS2 is to provide technical and functional specifications for the development of service interfaces based on the *Service Oriented Architecture* (SOA) that allow the management, search and reading of terminological content defined both locally and internationally. The CTS2 standard provides a list of terminology resources and, for each of them, a list of functional profiles has been defined.

3. Results

The proposed architecture is composed by three elements: a relational database, where all the ontologies and information about them will be stored (already developed and not part of this project); a CTS2 compliant Windows Communication Foundation (WCF) service whose interface exposes all the operation contracts expected by the CTS2 standard (already developed and not part of this project); a CTS2 compliant web application, client of the WCF service. This web platform will give the user the possibility to easily use the functionalities of the CTS2 compliant Service, for example: create, update, delete a terminology resource; read a terminology resource in a specific context; search the terminology resources with specific features in a certain context and get the list of all the changes applied to a particular terminology resource. Moreover, the use of controlled vocabularies, defined in different countries, will help the data extraction and integration from different sources containing notes, reports and referrals written in the original languages of the countries involved in the present project.

4. Discussion & Conclusions

The proposed architecture allows the correct management of mappings between different ontologies over the time and to monitor the evolution of a terminology resource through different versions. Future works will be the development of the web application cited above and the possibility of using a non-SQL database instead of a relational one.

Acknowledgements

The presented work has been financed by The Virtual Brain Cloud project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826421.

References

- [1] H. Min *et al.*, "Towards a standard ontology metadata model," in *CEUR Workshop Proceedings*, 2016.
- [2] C. González, B. G. M. E. Blobel, and D. M. López, "Formal specification of an ontology-based service for EHR interoperability," in *Studies in Health Technology and Informatics*, 2012.
- [3] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," *Int. J. Hum. - Comput. Stud.*, vol. 43, no. 5–6, pp. 907–928, 1995.
- [4] R. Gazzarata *et al.*, "A Terminology Service Compliant to CTS2 to Manage Semantics within the Regional HIE," *Eur. J. Biomed. Informatics*, 2017.

Providing Personal Health Records Based on a Mobile Application for Emergencies

Su Min KIM^a, Suyoung YOO^a, Won Chul CHA^{a,b}, Tae Rim KIM^{a,b}

^a*Samsung Advanced Institute for Health Science & Technology (SAIHST),*

Sungkyunkwan University, Seoul

^b*Samsung Medical Center, Seoul*

Abstract. We developed the service showing patient health record altogether which is managed by each hospital separately and recording the patient health information based on mobile application. We evaluated the effectiveness and satisfaction of the service. This study is aimed to reduce instances where patient's medical records are unknown to the medical staff in emergencies and allow patients to use and utilize their health information.

Keywords. Personal medical information, personal health report, mobile application, emergency

1. Introduction

With the aging and increase of chronic diseases, the importance of continuous health care is increasing. In addition, there has been an increasing demand for patients to manage and use their own medical information previously held by medical institutions. However, there was a limit to developing an appropriate platform between information providers and consumers. The purpose of this study is to develop appropriate channels for returning sovereignty of medical information to patients.

2. Method

The three hospitals, including Samsung Medical Center, provided application-based services for emergency patients called Firster to 1,000 emergency patients or carers and 12 medical staffs. Patients and carers are surveyed on the usability and satisfaction and the System Usability Test (SUS). Set a target of 69.5, which is slightly above the 68 SUS score that 50% of people in the reference evaluate, a total sample size of 1,000 was obtained based on one sample and t test.

Afterwards, interviews to Patients, carers and medical staffs are conducted to further explore usability and satisfaction.

3. Results

This study was conducted from October to November 2019 and the data is analyzing, thus this paper included only a few results of the data analysis. A total of 1,000 patients or carers participated in the demonstration, and the average of system usability test (SUS) was 67.08 (SD=13.8), which was lower than the average of 68 as the reference.

However, The lower the age, the higher the usability and satisfaction.

As a result of dividing into two groups based on the average SUS score of 68, the age group under 56 showed a high rate in the group with over SUS 68. As a result of comparing the usability of the smartphone and the usability of the app, the group with over SUS 68 showed higher usability than the group with under SUS 68.

Table 1 Comparison between two groups by SUS score under and over 68

	Total (N=1,000,100%)				Under SUS score 68 (N=560, 56.0%)		Over SUS score 68 (N=440, 44.0%)	
	age	n	%	SUS	n	%	n	%
	19~35	374	37.4	69.6(12.3)	178	31.8	196	44.5
	36~55	452	45.2	68.0(13.9)	252	45.0	200	45.5
	Over 56	174	17.4	59.5(14.7)	130	23.2	44	10.0
					mean		mean	p-value
Smartphone usability*					3.96(sd=0.61)		4.2(sd=0.46)	<0.05
SUS*					57.6(sd=9.82)		79.1(sd=7.40)	<0.05

* Smartphone usability: 5 score is the highest

4. Discussion and conclusion

The average of SUS was 67.08 (SD=13.8), which was lower than the average of 68 as the reference. But the period for demonstration was just 2 months and it was conducted only three hospitals. Therefore, we cannot predict the service usability and satisfaction of this service because it is still in the process of proof. Through demonstration, the usability of the app service varies with age and smartphone using status; The older people tend to have lower usability than the younger people do. There should be efforts to improve the usability of the high-age population.

It is expected that the personal health information utilization service will improve access to health information that has been managed separately by each medical institution and facilitate the utilization of health information, thereby reducing emergency difficulties and responding to the needs of patients for health information.

References

[1] James S. Kahn, Venu Aulakh, and Adam Bosworth, What It Takes: Characteristics Of The Ideal Personal Health Record, VOL. 28, NO. 2: Stimulating Health It, 2009

[2] Charalampos Doukas,Thomas Pliakas and Ilias Maglogiannis, Mobile Healthcare Information Management utilizing Cloud Computing and Android OS, 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology

[3] Michael Kyazze, Janet Wesson And Kevin Naude, The design and implementation of a ubiquitous personal health record system for South Africa, Stud Health Technol Inform. 2014;206:29-41

[4] Nidhi Bouri, MPH; Sanjana Ravi, MPH. Going Mobile: How Mobile Personal Health Records Can Improve Health Care During Emergencies. JMIR Mhealth Uhealth 2014;2(1):e8

[5] Jeongeun Kim, RN, PhD; Hongju Jung; David W. Bates, MD, MSc. : History and Trends of “Personal Health Record” Research in PubMed. HIR 2011 March;17(1):3-17

[6] Aaron Bangor, PhD, CHFP; Philip Kortum, PhD; James Miller, PhD. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies Vol. 4,- 3, May 2009:114-123

Qualitative Study to Design an Online Community for Patients with Psoriasis

Giuliana COLUSSI^{1a}, Janine SOMMER^a, Mariana SIMÓN^a, Lucila BRUCHANSKI^a,
Daniel LUNA^a

^a Health Informatics Department, Hospital Italiano de Buenos Aires, Argentina

Abstract. The aim of this research was to understand the needs of patients with psoriasis and the concerns of specialists in this pathology. Interviews with adults with psoriasis and dermatologists were conducted. We found 4 main dimensions: Frequently asked questions; Social stigma; Education to patients and Patient empowerment. The findings represent a first approach to understand the needs and perceptions of patients and dermatologists, for the design of a virtual community.

Keywords. Self-Help Groups, personal health records, psoriasis

1. Introduction

A virtual community in healthcare refers to a group of people who interact with each other, performing activities related to healthcare and education [1]. Patients with a recent diagnosis in which doubts and fear may arise, could find in virtual communities psychological support [2] by contacting people who have gone through similar circumstances. Participation of patients with long-term health conditions as psoriasis, improves self-control of the disease, produces positive outcomes related to health [3], facilitates shared decision making [4] and can even reduce mortality [5]. The objective of this research was to understand the needs of patients with psoriasis and the concerns of specialists in this pathology; for the design of a virtual community of patients with psoriasis into our PHR, applying user-centered design techniques.

2. Methods

We conducted a qualitative study, based on in-depth, semi-structured individual interviews to both patients and physicians. Patient selection was based on a convenience sample of adult patients diagnosed with psoriasis, whereas for physicians selection, we included dermatologists specialized in this disease. Each interview was audio-recorded, with the prior informed consent and protecting participants' confidentiality. We proceeded to analyze and categorize the information in domains according to the main findings. The analysis was performed through codification and categorization of the information, based on a process of constant comparison [6].

¹ Corresponding Author: E-mail: giuliana.colussi@hospitalitaliano.org.ar.

3. Results

We conducted 15 interviews with patients and 9 with dermatologists. We found 4 main dimensions. 1) *Frequently asked questions*: Patients were frequently interested on finding out causes of their disease, types of treatments available and treatment options that have shown better results. 2) *Social stigma*: Beyond the typical physical signs of the disease, the major source of patient discomfort came from social stigma and discrimination. An informed patient, might be beneficial for the community by promoting the dissemination of information on the the non-contagious status of the disease and thus, helping to overcome barriers in terms of social stigma. 3) *Patient Education*: Most patients will have read information on the internet, by the time they arrived to the physician appointment. Dermatologists argued about the difficulties they face trying to change patients minds once they have already acquired false beliefs and knowledge. 4) *Patient empowerment*: Most participants mentioned that it would be interesting to include questionnaires about the longitudinal follow-up of this pathology. and it would be interesting for patients to self-report these questionnaires in order to have a complete follow-up. Besides, it could represent a benefit for doctor-patient relationship since the time they now-a-days expend completing these questionnaires during the office visit, could be used to address other important topics.

4. Discussion

The main findings obtained from the interviews were related to information and education needs along with recommendations regarding psoriasis management, as well as the physical, emotional and social impact generated on people and the approach from the patient and family perspective [7].

We detected the need of a technological tool capable to give patients access to reliable information sources, promote online information sharing with the medical team or peers, and facilitate data management according to users experience [7]. The main findings will help us to develop a tool focused on user needs. Conclusively, this research represents the first step in the design of a virtual community by understanding the needs and perceptions of patients and dermatologists.

References

- [1] S. Ziebland, and S. Wyke, Health and Illness in a Connected World: How Might Sharing Experiences on the Internet Affect People's Health?, *Milbank Q.* **90** (2012), 219–249.
- [2] P.K.H. Mo, and N.S. Coulson, Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS, *Psychol. Health.* **27** (2012) 445–459.
- [3] L.F. Pendry, and J. Salvatore, Individual and social benefits of online discussion forums, *Comput. Human Behav.* **50** (2015), 211–220.
- [4] Y.K. Bartlett, and N.S. Coulson, An investigation into the empowerment effects of using online support groups and how this affects health professional/patient communication, *Patient Educ. Couns.* **83** (2011), 113–119.
- [5] W.R. Hobbs, M. Burke, N.A. Christakis, and J.H. Fowler, Online social integration is associated with reduced mortality risk, *Proc. Natl. Acad. Sci. U. S. A.* **113** (2016), 12980–12984.
- [6] R.H. Sampieri, C.F. Collado, and P.B. Lucio, *Metodología de la investigación*, McGraw-Hill Companies, 2006.
- [7] S.Z. Idriss, J.C. Kvedar, and A.J. Watson, The Role of Online Support Communities, *Arch. Dermatol.* **145** (2009).

Results of the Use of the Teleconsultation Platform After 2 Months of Implementation

Emiliano LOPEZ^a Maia BERLIN^a Romina STEIN^a Erica COZZI^a Andrea BERMUDEZ^a Humberto MANDIROLA BRIEUX^a Martín DÍAZ MAFFINI^a Roberto MOLDES^b Teodorico BOUSQUET^b Patricia MAC CULLOCH^b Daniel A. RIZZATO LEDE^a Cintia D SPERANZA^a and Alejandro LÓPEZ OSORNIO^a

^aMinistry of Health, Argentina

^bMinistry of Modernisation, Argentina

Abstract. The Ministry of Health (MoH) set the National Digital Health Strategy 2018-2024 as a state policy. It included a National TeleHealth Plan to enhance access and quality of healthcare in a wide territory like Argentina, leveraging more than 20 years of national telemedicine experiences and coordinating it with the territorial integrated health service networks proposed by the Universal Health Coverage strategy. In collaboration with the Ministry of Modernisation, the MoH developed and implemented a new TeleHealth Web Platform to perform eReferrals and eConsultations nationwide. This poster describes the first 2 months of usage.

Keywords. Telehealth, eConsult, eHealth, Telemedicine.

1. Introduction

Digital Health is one of the three pillars for the effective implementation of Universal Health Coverage in Argentina. The Ministry of Health (MoH) created a National Directorate of Health Information System to coordinate the National Digital Health Strategy 2018-2024. It included a National TeleHealth Plan to promote the development of this technology.

In collaboration with the Ministry of Modernisation, the MoH developed and implemented a new TeleHealth Web Platform to perform eReferrals and eConsultations nationwide. The teleconsultation platform allows primary care physicians located in remote areas to contact specialist for difficult patient cases[1], improving communication, and eventually avoiding physical referral. It was designed to close territorial specialist availability gaps[2]. In-advance training and education for leaders and end-users is one of the critical factors[3]. Governance and management aspects of Telemedicine are crucial, and have the potential of transformative challenge for the health system [4]. The platform was developed using international standards suggested by the MoH. The components were: the front end; a layer of data services; a back end that is responsible for persisting and retrieving information and a server to store the database. We used Fast Healthcare Interoperability Resources (FHIR), such as service request. This poster explains the first overview of its implementation.

2. Materials and Methods

The analysis of the implementation was done using de-identified data exported from the system, regarding institutions, specialties, telemedicine programs and teleconsultations.

3. Results

The Platform implementation was rolled out in 2 phases: an initial pilot with 2 provinces and a big bang adoption in the whole country afterwards. Previously, main leaders took a telehealth course provided by Mina Gerais University. Final users were trained locally. Explanatory videos and manuals were available through the MoH website. Initial parameterization, helpdesk and support was in charge of MoH personnel. Telemedicine Programs were created based on specialties and health disorders coded with SNOMED. Each province created healthcare institution networks, called Subsystems, and set the rules for the teleconsultations workflow. Interestingly, 23 out of 24 (95.8%) provinces started using the teleconsultation platform.

During the first 2 months, 18 telemedicine programs were created. The most used were Tele-Pediatrics (TPed) and Tele-Internal Medicine (TIM). The TPed program is well consolidated, based on a longstanding experience since 1987. Even though in August 2019 the consultations were 87.14% for TPed and 12.86% for TIM, in September the adult program usage increased 146%: 81.1% (TPed); 18.6% (TIM).

Regarding the reason for teleconsultation, 31.8% were initiated to reach a diagnosis, 20.4% were looking for treatment advice, 18% for physical referral request. Few consultations asked for an interventional procedure (11.3%), while 10.93% needed further lab or image studies. Finally, other categories were counter-reference (7.03%), prevention (0.35%) and rehabilitation (0.18%).

4. Conclusion

The TeleHealth Web Platform for eConsultations is a promising tool to close accessibility gaps and enhance territorial integrated health service networks.

References

- [1] C. Liddy, I. Moroz, A. Mihan, N. Nawar, E. Keely. A systematic Review of Asynchronous, Provider-to-Provider, Electronic Consultation Services to Improve Access to Specialty Care Available Worldwide. *Telemedicine and E-Health*. **25**(2019) 184-198.
- [2] VG Vimalananda, et al. Electronic consultations (e-consults) to improve access to specialty care: A systematic review and narrative synthesis. *Journal of Telemedicine and Telecare*. **21**(2015) 323–330.
- [3] A.F. Dos Santos, H.J. Alves, J.T. Nogueira, et al. Telehealth distance education course in Latin America: analysis of an experience involving 15 countries. *Telemed J E Health*. **20**(2014) 736-41.
- [4] Y. Gu, J. Warren, M. Orr. The potentials and challenges of electronic referrals in transforming healthcare. *N Z Med J*. **127**(2014) 111-8.

Risk Factors for Chronic Diabetes Patients

Oleg METZKER^a, Kirill MAGOEV^a, Stanislav YANISHEVSKIY^b, Alexey YAKOVLEV^b and Georgy KOPANITSA^a

^a*ITMO University, Saint Petersburg, Russia*

^b*Almazov National Medical Research Centre, Saint Petersburg, Russia*

Abstract. Specific predictive models for diabetes polyneuropathy based on screening methods, for example Nerve conduction studies (NCS), can reach up to AUC 65.8 – 84.7 % for the conditional diagnosis of DPN in primary care. Prediction methods that utilize data from personal health records deal with large non-specific datasets with different prediction methods. It was demonstrated that the machine learning methods allow to achieve up to 0.7982 precision, 0.8152 recall, 0.8064 f1-score, 0.8261 accuracy, and 0.8988 AUC using the neural network classifier.

1. Introduction

Every third patient with diabetes suffers from diabetic polyneuropathy (DPN). This complication is a severe problem for a chronic patient because it can be accompanied by neuropathic pain and lead to a decrease in the quality of life. Neuropathic pain significantly disrupts sleep, negatively affects daily activity and life satisfaction. The earlier the diagnosis is found, the easier the disease can be managed. Several studies on the prediction of diabetes mellitus that utilize Linear regression (LR) developed predictive models based on records with up to 967 independent variables available from large electronic health record archives resulted in up to 0.80% Area Under The Curve Receiver Operating Characteristics (AUC-ROC) and enabling to formulate the most likely trajectory of comorbidities. Application of the ensemble models approach based on the naïve Bayes, Linear regression (LR), Instance-Based Learner, support vector machine (SVM), artificial neural networks (ANN), decision trees, and random forest (RF) help to increase the accuracy of diabetes prediction with the accuracy of up to 74%. The goal of this study is the implementation of machine learning methods for early risk identification of diabetes polyneuropathy based on structured electronic medical records.

2. Methods

The dataset contains a total number of 238590 laboratory records. Each record contains patient and episode identifiers, a timestamp, and a varying number of measured parameters, the total number of which is 31 (27 laboratory tests, retinopathy, nephropathy, age, and gender). The records include information 5846 diabetic patients. Diagnosis served as the source of information about the target class values. To identify subclasses within the class of patients with diabetes, the classification problem was solved. Each experiment ran in the setting of stratified 5-fold cross-validation (i.e., random 80% of patients were used for training and 20% for testing, target class ratios in

the folds were preserved). The AUC was calculated based on an average of 5 curves (one curve per fold in the setting of 5-fold cross-validation).

3. Results

There are two large clusters (#0 and #3) that are opposite to the percentage of patients with polyneuropathy, namely the smallest and the largest rate of polyneuropathy. Cluster # 0 has an increased rate of patients with polyneuropathy: 59 %. In cluster #3, on the contrary, the reduced number of patients with polyneuropathy is 22 %. However, these are both female clusters. In male clusters #1 and #2, the percentage is close to the average. Cluster #4 is slightly below average. In clusters 0, 1, and 2, patients are about the same age. The difference in cluster 1 is that there is an outlier on the left, and it is male in contrast to cluster 0. Cluster 3 patients are young men, younger than patients cluster 0, 1, and 2, and also male. Cluster 4 are young male patients, which explains their low percentage of polyneuropathy. Cluster 5 patients of different ages with a small portion of polyneuropathy are mostly men. It is evident that the female gender and the age of more than 50 years is a risk factor for the development of polyneuropathy in patients with diabetes mellitus. However, there is a particular group of young men (cluster 4) with the probability of polyneuropathy. As a result of the cluster analysis, several groups of patients were identified: younger men and women (clusters 4 and 3, the share of polyneuropathy is significantly lower), older men and women (clusters 1 and 2: men, 0: women, the share of polyneuropathy is significantly higher). The most significant features were those of neutrophils and glucose level in the blood and the urine. The correlation matrix shows the high correlation of MPW, RBC, HGB, and polyneuropathy. If we discuss the main features (indicators) affecting the development of polyneuropathy in diabetes, the presence of a decrease in the relative number of segmented neutrophils, which supports the postulate of the main contribution of monocyte cells in the process of inflammation development [13], as well as signs of unsatisfactory glycemic control manifested in glucoseuria and hyperglycemia, become clinically important events.

4. Discussion and conclusion

Clustering shows differences within patients with diabetes mellitus. Decision trees demonstrate pathophysiological mechanisms. It was demonstrated that inclusion of just two features, namely “nephropathy” and “retinopathy”, allows to increase the performance even further, achieving up to 0.7982 precision, 0.8152 recall, 0.8064 f1-score, 0.8261 accuracy, and 0.8988 AUC using the neural network classifier.

Acknowledgements

This work financially supported by the government of the Russian Federation through the ITMO fellowship and professorship program. This worked was supported by a Russian Fund for Basic research [18-37-20002](#).

Rosacea Patients Are at Higher Risk for Obstructive Sleep Apnea: Automated Retrospective Research

Peter L. ELKIN^{1,2}, Sarah MULLIN^{1,2}, Sylvester SAKILAY^{1,2}

¹*Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, SUNY, USA*

²*Western NY VA, Department of Veterans Affairs*

Abstract. Using big data science we employ NLP and a novel interface the BMI Investigator to answer clinically meaningful questions. The use case presented is the association between Rosacea and Obstructive Sleep Apnea.

Keywords. Big data science, information Retrieval, NLP, Rosacea, Obstructive Sleep Apnea

1. Introduction

Automated retrospective research is a goal of an EHR driven research program. Here with this application we can go from clinical observation to science in ten minutes using a large cohort driven observational database linked to codified clinical notes using SNOMED CT. In this study we show that patients with Rosacea are at significantly increased risk for Obstructive Sleep Apnea (OSA).

Rosacea is a chronic skin condition that typically affects the face.[2][3] It results in redness, pimples, swelling, and small and superficial dilated blood vessels.[2] Obstructive sleep apnea (OSA) is the most common type of sleep apnea and is characterized by repeated episodes of complete or partial obstructions of the upper airway during sleep, OSA is usually associated with a reduction in blood oxygen saturation.

The BMI Investigator is a computer human interface built in .Net which allows simultaneous query of structured data such as demographics, administrative codes, medications (coded in RxNorm), laboratory test results (coded in LOINC) and formerly unstructured data in clinical notes (coded in SNOMED CT). The SNOMED CT codes are stored using Berkley DB and NoSQL Elastic Search, and the structured data is stored in SQL using the OMOP / OHDSI format. We have integrated genomic data into the software platform and are working to integrate image features into the query engine.

2. Methods

We use the high throughput phenotyping – Natural Language Processor (HTP-NLP) developed at the University at Buffalo.[4]. The BMI Investigator application was written in .Net and was created using the user-centered design development method with a team of programmers, physicians, graduate students and faculty [5,6]. We tested the system on

a population of 212,343 patients in our outpatient practices at the UBMD practice plans age 18 and older. The data for this trial was from 2010 to 2015. The data used in the system was judged by the IRB to be IRB Exempt #587570. The system allows users to use Boolean logic and parentheses to construct their queries.

Genomic data is presented as gene abnormalities that are used in clinical medicine and polymorphisms that have been identified are stored in a separate set of tables and they are also used to match to our patients who are included in the precision oncology project [7]. In an exploratory analysis, which took five to ten minute, we entered four queries into the BMI Investigator (Rosacea, Rosacea and OSA, OSA, total eligible cases in the database) and used a Pearson Chi-Square test to compare the results.

3. Results

The dataset has 212,343 patients. 211,764 patients did not have Rosacea. 5443 patients had OSA without Rosacea. 580 patients had OSA with Rosacea and 51 patients had OSA with Rosacea. The chance of OSA without Rosacea was 5443/211,764 (2.6%). The chance of OSA with Rosacea was 51/580 (8.8%). The Pearson Chi-Square test showed a p-value of <0.001. The relative risk of OSA given Rosacea was 3.4 times those patients without Rosacea. The number needed to test was only 12 to get one positive case.

4. Conclusions

The BMI investigator is an evolving tool linked to observational EHR data that has the ability to facilitate rapid retrospective research which can be clinically relevant. In this case we have shown evidence that patients with Rosacea are at increased risk for Obstructive Sleep Apnea and with a low number needed to test (NNT) they should all be asked the screening questions for OSA and if positive they should be considered for polysomnography (a sleep study). Many more studies and quality and safety projects are possible using the EHR based data and the BMI Investigators.

References

- [1] Sand, M; Sand, D; Thrandorf, C; Paech, V; Altmeyer, P; Bechara, FG (4 June 2010). "[Cutaneous lesions of the nose](#)". *Head & Face Medicine*. **6**: 7. doi:10.1186/1746-160X-6-7. PMC 2903548. PMID 20525327.
- [2] Tüzün Y, Wolf R, Kutlubay Z, Karakuş O, Engin B (February 2014). "Rosacea and rhinophyma". *Clinics in Dermatology*. **32** (1): 35–46. doi:10.1016/j.clindermatol.2013.05.024. PMID 24314376.
- [3] van Zuuren, EJ; Fedorowicz, Z (September 2015). "Interventions for rosacea: abridged updated Cochrane systematic review including GRADE assessments". *The British Journal of Dermatology*. **173** (3): 651–62. doi:10.1111/bjd.13956. PMID 26099423.
- [4] Schlegel DR, Crouner C, Lehoullier F, Elkin PL. HTP-NLP: A New NLP System for High Throughput Phenotyping. *Stud Health Technol Inform*. 2017;235:276-80.
- [5] Beuscart-Zephir MC, Elkin P, Pelayo S. Human factors engineering for clinical applications. *Stud Health Technol Inform*. 2006;124:685-90.
- [6] Husser CS, Buchhalter JR, Raffo OS, Shabo A, Brown SH, Lee KE, PL Elkin. Standardization of microarray and pharmacogenomics data. *Methods Mol Biol*. 2006;316:111-57.
- [7] Elkin PL, Mullin S, Sakilay S. *Biomedical Informatics Investigator*. *Stud Health Technol Inform*. 2018;255:195-199

Safety and Drugs: How Do We Record Medication Consumption and Prescription in Electronic Medical Records? A Look on Aspirin

Mercedes VOLPI^{a,1}, Santiago ESTEBAN^a and Sergio TERRASA^a

^a*Family and Community Medicine Division, Hospital Italiano,
Buenos Aires, Argentina*

Abstract. Background: documentation of aspirin consumption is usually not complete and clear, probably because it's a cheap drug that can be easily acquired. Objective: To evaluate the quality of the record of aspirin consumption documented in the electronic medical records (EMRs) of a Private University Hospital of Argentina. Design: Qualitative and quantitative descriptive exploratory study. Results: principal findings were that 86 % of the notes mentioned aspirin as a chronic prescription. 12% mentioned its temporary suspension. Numerous EMRs mentioned the use of an "antiplatelet" drug without specifying which, and didn't specify abbreviations, consumed dose or when to start or stop aspirin. Conclusions: overall quality of aspirin related information registration in EMRs was poor. This is concerning since it's a frequently prescribed drug, not exempt from adverse events.

Keywords. Aspirin, record, EMR, consumption

1. Background

Although medication consumption is the first cause of adverse events [1][2], information about drugs taken by patients is often not documented in electronic medical records (EMRs). For example, aspirin consumption (dosage, initiation or suspension) is not usually complete and clear, probably because it's a cheap drug that can be easily acquired over-the-counter.

2. Objective

To evaluate the quality of aspirin related consumption information documented in the EMR of a Private University Hospital of the Metropolitan Area of Buenos Aires, Argentina.

3. Design

We developed a keyword-based filter using expert knowledge, that allowed us to detect clinical notes in which aspirin was mentioned. We tested it on a manually annotated

¹ Corresponding author E-mail: mercedes.volpi@hospitalitaliano.org.ar.

sample of 2000 clinical notes, obtaining a sensitivity of 100% (95% CI 82.3 - 100) and a specificity of 100% (99.6 - 100). Using the filter, we obtained a sample of 1752 clinical notes which contained at least one of the keywords. These notes were manually reviewed and classified according to the information on aspirin consumption that they contained in the following categories: *current, chronic or temporary consumption of aspirin, definitive or temporary suspension, and indication to start it permanently or temporarily*. We also performed a qualitative description of how professionals document its use.

4. Results

86% (95% CI 84.2 - 87.6) of the notes referred to the current use of aspirin as a chronic prescription, 12% (10.5 - 13.6) mentioned its temporary suspension, 6.9% (5.7 - 8.2) mentioned the indication of starting it daily, 1.4% (0.91 to 2.1) referred to a temporary initiation and 1.2% (0.7 to 1.8) a permanent suspension. Numerous records did not clarify the dose the patient consumed or when they should start or stop taking it. We noted abbreviations of aspirin (i.e. omission of letters) that had not been clarified previously and many records mentioned the use of an "antiplatelet" drug without specifying it.

5. Discussion

Registration errors were more severe in notes describing changes of aspirin use. We did not review the structured systems that document the use of drugs in EMR. We only focused on unstructured clinical notes where health professionals often write changes in pharmacological treatments. We wonder what would be the closest to an ideal model of medication reconciliation and if a single health record compatible with the different levels of care would be possible.

6. Conclusions

Although there is consensus that having a complete pharmacotherapeutic history of patients optimizes health decision-making, we found a poor quality of registration of an accessible and inexpensive drug such as aspirin, which is not exempt from adverse events and whose chronic use is frequent.

References

- [1] Aibar-Remón C, Aranaz-Andrés JM, García-Montero JI, Mareca-Doñate R, Research on patient safety: needs and perspectives, *Med Clin (Barc)* **131** (2008), Suppl 3: 12-7.
- [2] Aranaz-Andrés JM, Limón R, Mira J.J., Aibar C, What makes hospitalized patients more vulnerable and increases their risk of experiencing an adverse event?, *International Journal for Quality in Health Care* **23** (2011), 705-712.

Service Personalization in IT-Supported Finnish Healthcare

Olli KORHONEN^{a,1}, Vasiliki MYLONOPOULOU^b, and Guido GIUNTI^a

^aUniversity of Oulu

^bUniversity of Gothenburg

Abstract. This paper reports a case study on the spontaneous personalization discussions emerged from interviews with healthcare professionals when asked about their work practices and the role of information technology (IT) during consultations. We thematically analyzed the personalization elements using an existing personalization framework to provide insights on the service personalization. Our results contribute to the better design of IT solutions that can support health services' personalization.

Keywords. Personalization, Healthcare, Service Design, Case Study

1. Introduction

As the Healthcare User's (HU) characteristics, medical condition, and personal preferences vary, they should be treated individually making Health Service personalization (HSP) important. HS often take place between the healthcare professional (HP) and HU, supported by IT [7]. IT can act as facilitator for HSP, providing data about HU's particular characteristics [2] to be used by HP and HS during HS delivery [5,7]. In service design, the users and their context are at its core [1]. This paper provides insights to the personalization field from a service design perspective [4] by presenting personalization elements, emerged by the analysis of an interview study [6] with a personalization framework [5].

2. Methods

We conducted 60 min semi-structured interviews with 6 HP working for more than 10 years in different levels in the Finnish primary care on behavioral change. It was a convenience sample of 1 nurse, 2 medical doctors, and 3 nutritionists. They described their practices, and the IT support in HS. The interview guide had 3 areas: HP background, work practices, and IT use. The guide did not include personalization-related questions. Authors independently thematically analyzed and later discussed the data related to personalization. To interpret our results, we used a personalization framework [5] related to the automation level IT provides for service personalization.

¹Corresponding Author, E-mail: olli.korhonen@oulu.fi.

3. Results

Coercive personalization refers to predetermined regulations influencing the HS personalization e.g. different IT-solutions offered in Finnish primary healthcare. Within the interviews, HP referred to the coercive personalization elements implicitly as the support systems and restrictions that were placed on their daily work.

Data display personalization refers to IT enabled personalized data for the stakeholders to interpret and make personalization related decisions. The HP used digital and none-digital types of data display support to illustrate cases of optimal behavior (e.g. a good meal portion image) which they discuss with the HU in comparison to their data. Data display support used also for progress monitoring during the behavioral change. The HU's (digital or paper-based) diary inputs lead to constructive discussions during consultation, helping HP to better understand and support their HU.

Collaboration-based personalization refers to IT supported personalization between stakeholders, HP and HU. In this case, it was mainly generated verbally during the consultation but counted with the support by technologies. HP and HU discussed to provide information and create understanding on HU's current and past behavior, context, life situation, motivation, and social circle.

4. Discussion & Conclusion

The need for personalized solutions is frequently presented in the literature [3]. We applied the service design approach in healthcare setting, which is uncommon [4] as quantitative approaches are more valued than qualitative [8]. The use of a personalization framework [5] resulted in a better understanding of the HP views, which may help improve IT solutions design. **Limitations.** The study conducted in Finland, the results may differ in a country with different culture and socioeconomic environment. The sample was small, but due to HP's different specialties and longtime experience, was considered representative of a primary healthcare setting.

References

- [1] Bruce, M. & Bessant, J.R. *Design in Business: Strategic Innovation Through Design*. Financial Times/Prentice Hall, 2002.
- [2] Fichman, R.G., Kohli, R., & Krishnan, R. Editorial overview—the role of information systems in healthcare: current research and future trends. *Information Systems Research*, **22** (2011), 419–428.
- [3] Giunti, G. 3MD for chronic conditions, a model for motivational mhealth design: embedded case study. *JMIR serious games*, **6**, (2018) e11631
- [4] Ingeborg, G., Melles, M., Stiggelbout, A., & Snelders, D. “The Potential of Service Design for Improving the Implementation of Shared Decision-Making.” *Design for Health* **1** (2017), 194–209.
- [5] Korhonen, O., & Isomursu, M. Identifying Personalization in a Care Pathway: A Single-case Study of a Finnish Healthcare Service Provider. In *Proceedings of the 25th European Conference on Information Systems* (2017), 828–841.
- [6] Mylonopoulou, V. Design for health behavior change supportive technology: healthcare professionals' perspective. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, (2018), 82–92.
- [7] Sacchi, L., Lanzola, G., Viani, N., & Quaglini, S. Personalization and patient involvement in decision support systems: current trends. *Yearbook of medical informatics*, **24** (2015), 106–118.
- [8] Sandelowski, M. “Whatever Happened to Qualitative Description?” *Research in Nursing & Health* **23** (2000), 334–340.

Serving Health Emergency Responders Through Online Learning – Findings from OpenWHO's Global User Metrics

Heini UTUNEN^a, Ngouille NDIAYE^a, Ursula ZHAO^a, Richelle GEORGE^a,
Melissa ATTIAS^a, Gaya GAMHEWAGE^a
^a*World Health Organization*

Abstract. This poster presents some preliminary findings of the OpenWHO.org platform's global use trends, in terms of the geographical distribution and occupational characteristics of its users. Assessment of user profiles is essential to measure the platform's impact, most notably related to the attainment of its core mission: the provision of life-saving knowledge worldwide. A quantitative study was conducted on the global metrics of OpenWHO's user statistics. Common user categories encompassed a wide range of professional bodies and occupations, both within public health and beyond, ranging from students and volunteers, to WHO staff, to members of international organizations and NGOs. Global tendencies in platform use confirm that the mission of OpenWHO, to provide timely, up-to-date and easy-to-understand lifesaving knowledge to healthcare workers based in-country and responding to outbreaks at the front line, is being met.

Keywords. Knowledge transfer, health emergencies, outbreaks, knowledge to action, online learning, knowledge resources, OpenWHO

1. Introduction

OpenWHO.org was launched in 2017 by the World Health Organization's (WHO) Health Emergencies Programme (WHE). The free, self-paced and open-access online learning platform facilitates the transfer of the latest and most-relevant scientific, technical and operational knowledge to frontline responders. Planning for the worst-case scenario, a pandemic, the platform is designed to host millions of users simultaneously and provide an easy-to-access platform for WHO technical guidance.

This study reveals some user profile characteristics that can be viewed against the global MOOC learner demographics. Veletsianos & Shepherdson (2016) highlight MOOC learners spanning over 194 countries with most users in North America and Europe and a majority being male, 20-40 years old, with a college degree or higher [1]. While OpenWHO user data does not collect educational background, the gross use by country and user profile seems to align with the global MOOC user profiles.

2. Method

Anonymous profile information of over 75 000 users was retrieved from the platform database in the format of spreadsheet. A quantitative analysis was conducted, primarily

focusing on user demographics and geographical presence, as well as affiliations given at the time of first enrollment. A separate analysis was conducted of the geographical information of users enrolled in the 12 most-visited courses, many of which were created in support of past or ongoing responses to global health emergencies.

3. Results

As of September 2019, the platform had 75 000 unique users and 130 000 course enrolments in more than 70 courses, including courses in all UN official languages and 15 other languages. User analysis reveals that OpenWHO is accessed by users in over 190 countries. Half of the platform's users (46%) are located in the five top countries: the United States of America, India, Portugal, Nigeria and Saudi Arabia.

Among the top 20 countries by users, four are experiencing ongoing WHO graded health emergencies – the Democratic Republic of the Congo, South Sudan, Pakistan and Nigeria – and five others are outbreak and epidemic-prone. The data on user location confirm that OpenWHO delivers lifesaving knowledge to those responding to health emergencies anywhere in the world. However, in the top countries list, a total of 11 countries are high-income countries, mainly from North America and Europe. User profiles are majority male and aged 20-40.

Results from the analysis of users' professional profiles highlight the platform's popularity among healthcare workers at the country level in low- and middle-income countries and countries with health events, while the student category is more popular in countries with no WHO-graded emergencies, such as the USA, India and Portugal.

4. Discussion

Real-time health emergencies and outbreaks appear to be driving the platform's popularity in its top countries. OpenWHO has also been successful in achieving yet another one of its aims, to make health-related technical knowledge more accessible to all, including any interested user, evidenced by the fact that the platform's materials are being utilized by staff across ministries of health, the UN, humanitarian organizations and NGOs, as well as healthcare workers, government personnel, volunteers and students. This also shows a keen interest in online learning from countries on all continents.

5. Conclusion

OpenWHO's user analytics suggest that the platform is reaching the world of health emergencies, as per its founding principle. Further analysis of user metrics will follow to understand the use of various languages on the platform and how use may be impacted by different types of health emergencies.

References

- [1] Veletsianos G. & Shepherdson P. (2016) A systematic analysis and synthesis on the empirical MOOC literature 2013-2015. *International Review of Research in Open and Distributed Learning*, 17(2), pp. 1-24

Smartphone Applications for Range of Motion Measurement in Clinical Practice: A Systematic Review

Anne-Violette BRUYNEEL^{a,1}

^a*Department of Physiotherapy, School of Health Sciences, HES-SO/University of Applied Sciences and Arts Western Switzerland, Geneva, Switzerland*

Abstract. The aim of this study was to investigate the validity and reliability of range of motion measurement via smartphone applications. This literature review included 26 articles after the selection process. The validity and reliability analysis showed good mean results (ICC or $r > 0.83$). Thus, in clinical practice, photographic, goniometric and inclinometric smartphone applications can be used to measure joint angle, but with caution for cervical, hip and shoulder motions.

Keywords. Range of motion – application – smartphone - validity – reliability

1. Introduction

Many smartphone applications have been launched in recent years for patient management in clinical practice, as well as range of motion tools. However, before generalizing these tools, it is necessary to know the psychometrics qualities of these new smartphone applications. *The aim* of this study was to provide a systematic review about validity and reliability (intra and inter-raters) of range of motion measurement (upper-limbs, lower limbs and trunk movements) via smartphone application.

2. Study

2.1. Method

The research was conducted, in English, using the following keywords such as « healthy subjects », « phone », « range of motion », validity », « reliability » et « repeatability » on the databases: PubMed/ Medline, ScienceDirect and Pedro. A synthesis was carried out and the qualities of each study was evaluated by the Qare score.

2.2. Results

563 articles about smartphone applications used in healthy subjects were identified. Following the selection criteria, 26 articles were included which are six articles for cervical, thoracic and lumbar motions, 11 for lower limb and 9 for upper limb motions.

¹ Corresponding Author, E-mail: anne-violette.bruyneel@hesge.ch

These studies included 818 healthy subjects. The mean Qarel score for all studies was 7.36/12 (min:6/max:9). The synthesis showed that the for validity was good ($r = 0.83$ [min:0.40, max:0.99]). The intra-rater reliability analysis showed an ICC=0.88 (min:0.05 – max:0.99), investigated by 23 studies. The inter-rater reliability was assessed by 20 studies with a mean ICC result (ICC=0.85 [min:0.07 – max:0.99]). The results highlighted no difference between the type of applications (inclinometer or goniometer) and the phone system (iPhone vs. Android).

2.3. Discussion

This systematic review showed the existence of numerous studies on the psychometric qualities of smartphone applications to measure range of motion in clinical practice for healthy subjects. All studies presented a consistent methodology with the stated objective and tested validity prior to reliability. The data suggest that smartphone with the angle apps is a viable substitute for using a usual goniometer or an inclinometer when measuring angular changes that typically occur when examining range of motion and demonstrate the capacity of multiple examiners to accurately use smartphone-based angle apps. The trunk, upper limb and lower limb range of motion were assessed for validity and intra-rater reliability in healthy subjects. The inter-rater reliability has not been studied for hip movements. The best values were obtained for the wrist, the elbow, the ankle, the knee, the lumbar and thoracic range of motion, whatever application was used. No study at the time of the review was found on the finger movements.

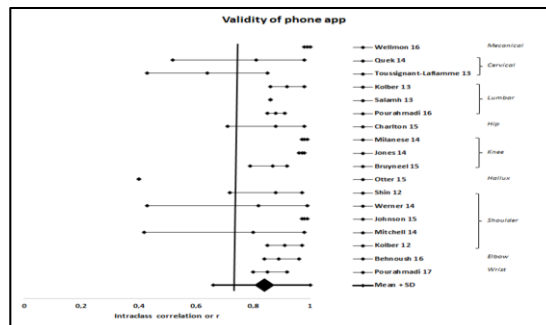


Figure 1. Synthesis of validity results.

References

- [1] Bruyneel AV. Validity and reliability of phone applications for range of motion measurement in clinical practice: Literature review. *Kinésithérapie, la revue* 2018 ;879:13-24.
- [2] Lee D, Han S. Validation of Joint Position Sense of Dorsi-Plantar Flexion of Ankle Measurements Using a Smartphone. *Health Inform Res*. 2017 Jul;23(3):183-188. doi: 10.4258/hir.2017.23.3.183.
- [3] Wellmon RH, Gulick DT, Paterson ML, Gulick CN. Validity and Reliability of 2 Goniometric Mobile Apps: Device, Application, and Examiner Factors. *J Sport Rehabil* 2016;25:371-9. doi:10.1123/jsr.2015-0041.

Social Media Mining for Postpartum Depression Prediction

Alina TRIFAN^{a,1}, Dave SEMERARO^b, Justin DRAKE^b, Radek BUKOWSKI^c and José Luís OLIVEIRA^a

^aDETI/IEETA, University of Aveiro, Portugal

^bTexas Advanced Computing Center, University of Texas at Austin, USA

^cDell Medical School, University of Texas at Austin, USA

Abstract. This study investigated the feasibility of a postpartum depression predictor based on social media writings. The current broad use of social media networks generates a large amount of digital data, which, when coupled with artificial intelligence methods, have the potential to disclose significant health related insights. In this paper we explore the use of machine learning for prediction of postpartum depression on a corpus created from Reddit posts.

Keywords. postpartum depression, social media, machine learning, mental health, women's health

1. Introduction

Despite the widespread use of prenatal care and the increasing adoption of ante- and post-partum depression screening tools, post partum depression (PPD) often goes undiagnosed or untreated. For example, Ko et. al [1] estimated nearly 50% of of pregnant women who experienced a major depressive episode in the past year received treatment, and Vesga-Lopez et al. [2] estimated the prevalence for untreated, postpartum mood disorders to be 85%. Clearly, there is an opportunity to enhance existing screening tools and protocols to detect PPD, and to promote better treatments. Here, this study considers the use of social media as an early intervention tool by accelerating the potential diagnosis of mothers at risk of PPD and thus the implementation of early interventions.

2. Methods

We build a corpus of postpartum depression related posts matched with control posts using the Reddit API². We collected public posts from PPD related subreddits, along with posts from parenting subreddits as controls. We conducted this study as a binary classification question. Prior to post classification, the preprocessing of the Reddit posts followed standard approaches in text classification. The posts were lowercased and tokenized, after removing all non-alphabetic characters. Stopwords were filtered. We considered Bag of Words (BoW) features as predictors and we applied counts and tf-idf based feature weighting. All experiments were managed using the scikit-learn machine learning framework [3]. Additionally, our experiment targeted the relation between gen-

¹Corresponding Author: alina.trifan@ua.pt

²<https://www.reddit.com/dev/api/>

eralized depression and postpartum depression. For this purpose, we employed a model previously trained on a dataset containing writings of depressed Reddit users, matched with control users, the Reddit Self-reported Depression Dataset (RSDD) proposed by Yates et al. [4].

3. Results

We performed experiments with two different estimators, under different configurations, Passive Aggressive and Perceptron, as well as the model trained on RSDD. We present the best results in Table 1. These results demonstrate that it is possible to distinguish postpartum depression related context from generalized user writings with quite a high degree of confidence. Moreover, as expected, there is a clear connection between the writing style of people suffering from depression, independently of their age and sex, and recent mothers that go through postpartum depression.

Table 1. Comparative results on classifying postpartum depression.

Method	Prec.	Rec.	F1	Acc
Perceptron	0.83	0.87	0.85	0.89
Passive Aggressive (loss=sqrt_hinge)	0.90	0.89	0.9	0.93
RSDD trained model	0.95	0.43	0.6	0.8

4. Conclusions

The preliminary results that we present in this paper serve as a motivation for safely exploring social media networks as enablers of early risk interventions for mothers suffering from postpartum depression. As future work, we will focus on extending the collected corpus with more user histories and identifying psycholinguistics features that differentiate the postpartum depression group from the control group within our corpus.

Acknowledgements

This work was supported by the Integrated Programme of SR&TD SOCA (Ref. CENTRO-01-0145-FEDER-000010), co-funded by Centro 2020 program, Portugal 2020, European Union, through the European Regional Development Fund and by UTAustin Portugal through the Advanced Computing Training Program.

References

[1] Jean Y. Ko, Sherry L. Farr, Patricia M. Dietz, and Cheryl L. Robbins. Depression and Treatment Among U.S. Pregnant and Nonpregnant Women of Reproductive Age, 2005–2009. *Journal of Women’s Health*, 21(8):830–836, June 2012.

[2] Oriana Vesga-López, Carlos Blanco, Katherine Keyes, Mark Olfson, Bridget F. Grant, and Deborah S. Hasin. Psychiatric Disorders in Pregnant and Postpartum Women in the United States. *Archives of General Psychiatry*, 65(7):805–815, July 2008.

[3] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[4] Andrew Yates, Arman Cohan, and Nazli Goharian. Depression and self-harm risk assessment in online forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, page 2968–2978. Association for Computational Linguistics, 2017.

Specification and Distribution of Vocabularies Among Consortial Partners

Mirko GRUHL^{a,1}, Ines REINECKE^a and Martin SEDLMAYR^a

^aCenter for Medical Informatics, Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany

Abstract. Due to the variety of different software systems and disparate observational databases, the need for a uniform data representation rises. Common data models (CDM) support the harmonisation of data. A powerful but compact software setup and a minimum vocabulary set has been composed via Docker to facilitate analysis of data across ten university hospitals. The presented approach also creates the possibility to use a concise database which is easy to deploy.

Keywords. vocabulary, OMOP, Docker, reproducible, data sharing

1. Introduction

To overcome the issues caused by heterogeneous databases, common data models (CDM) support the harmonisation of data. As part of the MIRACUM initiative [1], all data should be transformed into the Observational Medical Outcomes Partnership (OMOP) CDM [2]. This yields a universal format and a corporate representation with corresponding terminologies, vocabularies and coding schemes. The outcome of the implementation is the utilisation and analysis of data across ten university hospitals and the possibility to use the OHDSI suite of applications and exploration tools [3] in one ready-made distribution.

2. Method

First, a minimal vocabulary for a provided synthetic test data was specified by downloading the full set of standardised vocabularies and based on local data and mappings the actually needed concepts were determined. The next step comprised the enhancement of the minimal vocabulary to meet the requirements of the core data. Based on the work of Maier et al. [4] who integrated the ICD-10-GM (International Classification of Diseases, German Modification) and the "Operationen- und Prozedurenschlüssel" (surgery and procedure key) the vocabulary set was cleaned from unnecessary vocabularies that have no relevance in Germany. Finally, a system was planned and realised

¹Corresponding Author: E-mail: mirko.gruhl@uniklinikum-dresden.de

using Docker. The software bundle includes the OHDSI WebAPI [3] which contains all services that can be called from Atlas [3]. Additionally, the R Methods Library (<https://github.com/OHDSI/Broadsea-MethodsLibrary>) and Achilles [3] were applied. The databases were generated reproducibly via scripts and shared for the partners as Docker volumes which differ in the setup and size to meet different requirements.

3. Results

All system dependent volumes and images are hosted in a common registry and self-hosted cloud storage, respectively. The developed system uses a Docker compose file which is derived from the OHDSI Broadsea [5] tool set. Thus, the whole setup can be deployed in minutes using simple docker commands. The project is publicly available via GitLab (<https://gitlab.miracum.org/gruhlmi/ohdsi-omop-v5>).

4. Discussion

The system can be reproducibly deployed. Broadsea was chosen as a basis and customized so that a complete database with all needed content to start work with OMOP is available from one hand. The OMOP CDM was selected because of initial experiences and a big community. The compilation of the provided minimum vocabulary fits the requirements for healthcare research in Germany. Hence, every consortial member can start up an initial database with optimal memory utilisation. The use of the vocabulary is subjected to the license agreement and expects individual responsibility by every end user. Apart from the promising approach there are still challenges regarding the particularities of German classifications and procedures and the mapping to standard vocabularies.

5. Conclusion

The successful realisation of the objectives yields a powerful system which uses OMOP as starting point for federated healthcare research among ten MIRACUM sites. Further work on this topic will include the upgrade of the components and a strategy development to update the vocabularies automatically. The work is part of the project MIRACUM, funded by the German Ministry of Education and Research (FKZ 01ZZ1801L).

References

- [1] Prokosch, Hans-Ulrich, et al. "MIRACUM: medical informatics in research and care in University medicine." *Methods of information in medicine* 57.S 01 (2018): e82-e91.
- [2] Overhage, J. Marc, et al. "Validation of a common data model for active safety surveillance research." *Journal of the American Medical Informatics Association* 19.1 (2011): 54-60.
- [3] Hripcsak, George, et al. "Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers." *Studies in health technology and informatics* 216 (2015): 574.
- [4] Maier, C., et al. "Towards implementation of OMOP in a German university hospital consortium." *Applied clinical informatics* 9.01 (2018): 054-061.
- [5] OHDSI Broadsea. [cited 2019 Oct 01]. Available from: <https://github.com/OHDSI/Broadsea>

Standardized Communication Using FHIR and SNOMED CT in Treatment of Diabetic Foot Syndrome Within the Project iFoot

Dominik LUDMANN^{a,1}, Elisabeth PANTAZOGLOU^a and Hubert OTTEN^a
^a*Hochschule Niederrhein, University of Applied Sciences, Faculty of Healthcare*

Keywords. Interoperability, Standardization, Diabetic Foot Syndrome, FHIR, SNOMED CT

1. Introduction

About 15% of patients with diabetes mellitus develop a diabetic foot syndrome (DFS) during their lifetime. Complications caused by DFS are the main reason for hospital admissions and amputations [1] [2]. The number of amputations in DFS patients is estimated at 40,000 per year and accounts for approximately 70% of all amputations performed in Germany. The total costs of treatment in Germany amount to approximately 2.5 billion euros per year. In order to reduce expenses, the DFS needs optimal care and treatment [3]. The data and information transfer is a decisive aspect in ensuring the quality of care and patient safety. The development and use of modern e-health applications is therefore a promising approach for the interdisciplinary treatment of DFS. The iFoot research project focuses on an optimized approach to DFS medical and nursing care. In 2005 the German Federal Association for Statutory Health Insurance Physicians (Kassenärztliche Bundesvereinigung, KBV) introduced a specification for the electronic Disease Management Programme (eDMP) for patients with Diabetes mellitus type 1 and type 2. The aim of the eDMP is the structured treatment of chronically ill patients with the help of defined care processes based on defined therapy goals and to improve communication and care between service providers and the inpatient and outpatient sectors of the German health care system [4].

2. Methods

Since the current specification of the eDMP Diabetes mellitus type 1 and type 2 does not provide the use of terminology standards, some excerpts of the eDMP are represented in this paper using the internationally recognized IT standards Fast Healthcare Interoperability Resources (FHIR) and the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT). FHIR is an international standard of Health Level Seven (HL7) for the exchange of digital health data [5]. SNOMED CT is a multilingual medical terminology system that is already used in over fifty countries for the electronic representation of clinically relevant terms [6]. In the context of this ongoing work, an example will be used to illustrate the possibility of mapping the

¹ Corresponding Author, E-mail: dominik.ludmann@hsnr.de.

eDMP diabetes mellitus type 1 and type 2 according to the guidelines of the KBV with international standards.

3. Results

As an example, the mapping of the specified data fields for *further risks for ulcer* is carried out in an FHIR-Observation resource via SNOMED CT code *714664001 (At Risk of diabetic foot ulcer)*. Within the *Observation.valueCodeableConcept* element, it is possible to define the given values in coded form. For this purpose, the KBV defines the following seven concepts (Table 1):

Table 1. Mapping examples from eDMP Diabetes mellitus to SNOMED CT

Code	Term	Code System
373066001	Yes	SNOMED CT
373067005	No	SNOMED CT
229844004	Deformity of foot	SNOMED CT
75134009	Not examined for	SNOMED CT
396228006:363713009=255345002	Hyperkeratosis with bleeding	SNOMED CT
128926000:255234002=371087003	State after ulcer	SNOMED CT
128926000:255234002=180030006	State after amputation	SNOMED CT

4. Discussion

The current results of this ongoing work demonstrate that the use of FHIR and SNOMED CT is well suited and feasible for the annotation of the eDMP Diabetes mellitus type 1 and type 2 specification, although the annotation of the dataset poses some challenges due to a lack of information from KBV. Furthermore, as shown in Table 1, for more complex data elements (e.g. *Hyperkeratosis with bleeding*), post-coordinated SNOMED CT concepts had to be built. By using standardized terminologies, future implementations of the eDMP for Diabetes mellitus could reduce redundancies in data collection and storage. Thus, eDMPs could become an integral part of an interoperable eHealth landscape in the German health care system.

References

[1] Hochlenert, D., Engels. G., Morbach, S., Schliwa, S. & Game, F.L. Diabetic Foot Syndrome - From Entity to Therapy, Springer, Berlin-Heidelberg, 2018.

[2] Morbach, S., Müller, E., Reike, H., Risse, A., Rümenapf, G. & Spraul, M. Diabetisches Fußsyndrom, Georg Thieme Verlag KG, Stuttgart, New York, 2017.

[3] Deutsche Diabetes Gesellschaft, online, <https://www.deutsche-diabetes-gesellschaft.de/presse/ddg-pressemeldungen/meldungen-detailansicht/article/die-meisten-amputationen-in-deutschland-sind-folge-von-diabetes-und-liessen-sich-verhindern.html>, (accessed October 10, 2019).

[4] KV Telematik GmbH, eDMP, <https://www.kv-telematik.de/praxen-und-krankenhaeuser/informationen-zu-kv-connect/anwendungen/edmp/> (accessed October 10, 2019).

[5] HL7 Deutschland e.V., HL7 FHIR mobile Kommunikation und mehr <http://hl7.de/themen/hl7-fhir-mobile-kommunikation-und-mehr/warum-fhir/> (accessed October 10, 2019)

[6] SNOMED International, Wo we are, <http://www.snomed.org/snomed-international/who-we-are> (accessed October 10, 2019).

Strategy for the Analysis and Visualization of Electronic Medical Record Data for Public Hospitals in the City of Buenos Aires

Tomas GALLUZZI ^a, Marina RIDAO ^a, Santiago ESTEBAN ^a

^aHealth Information Management and Statistics Office, Ministry of Health, City of Buenos Aires, Argentina.

Abstract. Through the development of a dashboard with participative methodology we present a centralized strategy to analyze and visualize EMRs data for the management of 15 public hospitals from Buenos Aires City. This approach could constitute an efficient option for public health systems from developing countries.

Keywords. Electronic medical records, R software, data visualization, public health care.

1. Introduction

The availability and increasing amounts of data from electronic medical records (EMRs) give an important boost for information-based health service management. However, the exploitation of information to improve health care systems in low-income countries is scarce [1]. In June 2016, the Ministry of Health of the city of Buenos Aires began to implement an EMR system at the primary healthcare level and progressively expanded into general hospitals [2]. Nevertheless, hospitals do not have human resources with the skills to conduct data analysis from EMRs. To minimize these technical barriers and support decision-making processes, AGISE¹ designed a centralized and participative strategy to create dashboards for the management of 15 public hospitals.

2. Methods

The centralized component of the strategy included the analysis and development of the dashboards. R software was chosen for statistical computing as it is open-source and free tool. The following libraries were used for visualization: highcharter, to create charts, leaflet, to create interactive maps, and rmarkdown, to render the dashboards. AGISE scheduled meetings with the managers of the three hospitals with the highest levels of EMR implementation in order to define and design new indicators and review testing versions in a participative manner. Unstructured individual exchanges with each manager were conducted for indicator prioritization and selection.

¹ Health Information Management and Statistics Office (AGISE) whose objective is to provide timely and quality data from EMR through qualified professionals with heterogeneous profiles.

3. Results

After reaching a consensus on what indicators and information should be included, the final version was elaborated and then distributed to the remaining hospitals. The dashboard contains an initial page (Fig.1) divided into four sections: Demographic Data (Fig.2), Registration, Outpatient Consultations, and Emergency Room Consultations (Figs.3 and 4). New versions of the design of the dashboard are incorporated on an ongoing basis as more hospital areas implement EMR systems.

4. Discussion

Although management frameworks recommend decentralized decision making for hospital autonomy and governance [3], the centralized and participative approach may have advantages in a context of low-qualified human resources for data analysis.

5. Conclusion

This action plan could constitute a good option for decision-making processes in public hospitals using EMR information with available resources in health systems from developing countries. Next steps will include evaluation of dashboard use and utility.



Figure 1. Initial page for each hospital.

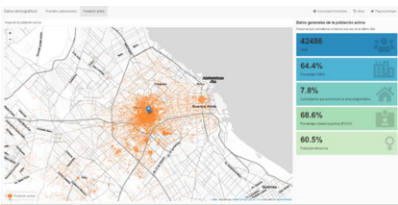


Figure 2. Demographic Data.



Figure 3. Outpatient Consultations.

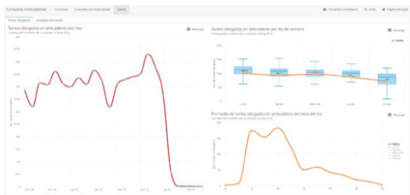


Figure 4. Outpatient Consultations.

References

[1] R. Wyber et al. Big data in global health: improving health in low- and middle-income countries. *Bulletin of the World Health Organization*. **93** (2015), 203-208.

[2] M.V. Giussi, A. Baum, F. Plazzotta, P. Muguerza, and F. González Bernaldo De Quirós, Change management strategies: Transforming a difficult implementation into a successful one, *Stud. Health Technol. Inform.* **245** (2017) 813–817.

[3] Rechel. B. Duran, A. Saltman R. What is the experience of decentralized hospital governance in Europe? 10 case studies from Western Europe on institutional and accountability arrangements. (Policy Brief, No. 28.) European Observatory on Health Systems and Policies, 2018.

Strengthening Behavior and Social Functioning Among Persons with Autism Spectrum Conditions Using Artificial Intelligence and Behavioral Activation: Protocol for the Well-Being and Health for Loved ones with ASD (WHOLE) Psychosocial Pilot Randomized Controlled Trial

Pierre BOWRIN^a and Usman IQBAL^{b,1}

^a *Global Health and Health Security Doctoral Program, College of Public Health, Taipei Medical University, Taipei, Taiwan*

^b *Global Health and Health Security, College of Public Health, Taipei Medical University, Taipei, Taiwan*

Abstract. Although a proportion of families and communities in low resource settings continue to provide care to loved ones with autism spectrum conditions, many of the affected persons remain undiagnosed and without access to proven therapies due to high treatment cost and cultural pressures, in particular. Use of conversational agents on mobile phones in combination with behavior activation home care may provide an innovative, culturally appropriate and affordable platform for strengthening behavior and social functioning outcomes, in addition to an opportunity for participation of the persons with autism spectrum conditions in the intervention development process. We aim to assess the effectiveness of an intervention that incorporates artificial intelligence conversational agent technologies and behavioral activation therapy techniques.

Keywords. Artificial Intelligence, Autism Spectrum Disorders, Conversational Agent, Chatbot, Psychosocial, Social functioning, Randomized Control Trial

1. Introduction

Autism Spectrum Disorders (ASD) are the fastest growing developmental disability [4, 5] that affects approx. 1% of the world population [2] and impacts behavioral and social functioning deficits such as social motivation, social anxiety, social cognition, and social skills [3].

¹ Corresponding Author, E-mail: usmaniqbal@tmu.edu.tw.

Autism policy and intervention research stakeholders advocate for psychosocial interventions that reduce core symptoms and improve adaptive skills; empower families and local communities; are applicable for low resource settings and actively incorporate persons with ASD in the research process and dissemination of findings [1], but interventions with such components remain underutilized. Therefore, this study aims to assess the feasibility of conducting a randomized control trial (RCT) evaluating the effectiveness of a behavior and social functioning intervention using artificial intelligence via a conversational agent or chatbot and behavioral activation with persons with Autism Spectrum Conditions.

2. Methods and Analysis

The Well-being and Health for Loved Ones with Autism Spectrum Conditions (WHOLE) intervention is a pilot, single-blinded, randomized controlled trial evaluated over 12 weeks, with 6 and 12-month follow-up periods to assess the feasibility, social and economic value of the intervention. Sixty youth/young persons will be recruited and randomized into intervention and control groups. Participants will be selected, in part, via the Emotional and Behavioral Screener instrument from the undiagnosed ASD population. Two separate pilot RCTs will be conducted in Saint Kitts & Nevis and Republic of China (Taiwan). Analysis of covariance (ANCOVA), descriptive statistics and social return on investment (SROI) analysis methods will be employed.

References

- [1] Amaral, D. G., Anderson, G. M., Bailey, A., Bernier, R., Bishop, S., Blatt, G., Canal-Bedia, R., Charman, T., Dawson, G., de Vries, P. J., Dickey-Bloom, E., Dissanayake, C., Kamio, Y., Kana, R., Khan, N. Z., Knoll, A., Kooy, F., Lainhart, J., Levitt, P., ... Whitehouse, A. (2019). Gaps in Current Autism Research: The Thoughts of the Autism Research Editorial Board and Associate Editors. *Autism Research: Official Journal of the International Society for Autism Research*, 12(5), 700–714.
- [2] De Luca, R., Leonardi, S., Portaro, S., Le Cause, M., De Domenico, C., Colucci, P. V., Pranio, F., Bramanti, P., & Calabrò, R. S. (2019). Innovative use of virtual reality in autism spectrum disorder: A case-study. *Applied Neuropsychology. Child*, 1–11.
- [3] Pallathra, A. A., Calkins, M. E., Parish-Morris, J., Maddox, B. B., Perez, L. S., Miller, J., Gur, R. C., Mandell, D. S., Schultz, R. T., & Brodtkin, E. S. (2018). Defining behavioral components of social functioning in adults with autism spectrum disorder as targets for treatment. *Autism Research: Official Journal of the International Society for Autism Research*, 11(3), 488–502.
- [4] Price, S. (2019). Addressing Autism: Giving Physicians Tools. *Texas Medicine*, 115(12), 42–44.
- [5] Thabtah, F. (2018). Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health & Social Care*, 1–20.

TEDIS, a Comprehensive Data Model for In-Depth Clinical Assessment of Patients Affected with Neuro-Developmental Disorders Including Autism

Mohamed BEN SAID ^{a,1}, Laurence ROBEL ^b, Pauline CHASTE ^c, Didier PERISSE ^d, Marie-Joelle OREVE ^e, Pascale ZYLBERBERG ^f, Anne PHILIPPE ^g, Catherine JOUSSELME ^h, Stéphanie LACOSTE ⁱ, Mario SPERANZA ^e, Ines HAFSA ^a, Fatouma CISSE ^a, Zoubair CHERQAOU ^a and Jean-Philippe JAIS ^a

^aParis Descartes University, Faculty of Medicine, Biostatistics Unit, Necker Hospital, APHP - Paris, France

^bDept of Child Psychiatry, Necker Hospital, APHP, Paris, France

^cDept of Child and Adolescent Psychiatry, Sainte Anne Hospital, Paris, France

^dDept of Child and Adolescent Psychiatry, GH Pitié-Salpêtrière, APHP, Paris, France

^eDept of Child and Adolescent Psychiatry, Versailles Hospital, Le Chesnay, France

^fUNIDEP, Jean Verdier Hospital – APHP - Bondy, France

^gDevelopmental Brain Disorders Laboratory, Imagine Institute, INSERM UMR 1163 – Paris Descartes University - Paris, France

^hFondation Vallée, Paris Sud University, Gentilly, France

ⁱDept of Child and Adolescent Psychiatry, GHEF Hospital – JOSSIGNY, France

Abstract. TEDIS, an information system dedicated to patients affected with neuro-developmental disorders including autism, focuses on patient data generated during in-depth clinical assessment in nine expert centers in Ile-de-France region. Long term partnership involving methodologists and domain experts is necessary to support quality data production and analyses and to guarantee quality data and information governance in a domain characterized by frequent evolutions in clinical assessment instruments and in diagnostic criteria and classification.

Keywords. Information model, data governance, information governance, DSM-5, ICD-11, Autism Spectrum Disorder, Neuro-Developmental Disorder.

1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by persistent difficulties in social communication and interaction, alongside with restricted, repetitive behavior patterns and interests [1]. The diagnosis “*is still made solely on the basis of self/informant reports of behavior according to diagnostic*

¹Corresponding Author, mohamed.bensaid@aphp.fr, mohamed.ben-said@ext.paris-descartes.fr.

criteria in the *Diagnostic and Statistical Manual of Mental Disorders [DSM-5]* or the *International Classification of Diseases [ICD-10]*” [2]. Psychiatric experts from the nine “Centres de Diagnostic et d’Evaluation de l’autisme” or (CDE) in Ile-de-France region, gathered with methodologists in a professional network to manage in a dedicated primary data source, valuable information resuming in-depth, multidisciplinary and psychopathological ASD’ patients’ assessment. The objectives are to improve knowledge about ASD disorders, follow-up patient cohorts and to improve patient care, decision making, and research. The process converged to the TEDIS project [3] supported by the regional public health agency, academic institutions and the regional ASD resources association.

2. Methods

An agreed-upon chart delimits the contours of the TEDIS project and defines policy for data information governance. An application system was developed to facilitate ASD’ patient data collection and quality control. Aggregated data are analyzed and described in an automatic feedback report [4]. Ethical authorizations facilitated Internet deployment via a Certified Health and Personal Data Internet Hosting Provider [5].

3. Results

An active expert network federated around ASD’ patient data production, analyses and maintenance. Up-to-date, 978 controlled ASD patient records in TEDIS’ cohort provide a quality data source supporting research and public health decision making.

4. Discussion / Conclusion

Challenges need being addressed to maintain the professional network active to capture domain expertise, to insure cohort progress and to keep the data model up-to-date with evolutions in diagnostic criteria (DSMIV/DSM-5) and classifications (ICD10/ICD11).

References

- [1] Autism: Recognition, Referral and Diagnosis of Children and Young People on the Autism Spectrum. National Collaborating Centre for Women’s and Children’s Health (UK). NICE clinical guidelines, No. 128. London: RCOG Press; 2011 Sep.
- [2] Ashwood KL, Buitelaar J, Murphy D, Sporeen W, Charman T. European clinical network: autism spectrum disorder assessments and patient characterisation. *Eur Child Adolesc Psychiatry*. 2015 Aug;24(8):985-95.
- [3] Ben Said M, Robel L, Golse B, Jais JP. Strengthening Data Confidentiality and Integrity Protection in the Context of a Multi-Centric Information System Dedicated to Autism Spectrum Disorder. *Stud Health Technol Inform*. 2017;245:1133-1137.
- [4] Xie Y. Dynamic Documents with R and Knitr, Second edition. Chapman & Hall/CRC The R series; Boca Raton , FL, 2016.
- [5] Labels et certifications - Hébergement des données de santé. [Cited on May 16th 2019] Paris – France. [Internet]: Available from: <https://esante.gouv.fr/labels-certifications/hebergement-des-donnees-de-sante>.

Telemedicine Interventions for the Management of Diabetes: A Systematic Review and Meta-Analysis

Stine HANGAARD^{a,1}, Sisse H. LAURSEN^b, Flemming W. UDSEN^a, Peter VESTERGAARD^c, and Ole HEJLESEN^a

^a*Department of Health Science and Technology, Aalborg University*

^b*University College of Northern Denmark, Department of Nursing*

^c*Steno Diabetes Center North Denmark*

Keywords. Telemedicine, diabetes, telemonitoring

1. Introduction

Telemedicine is often suggested as a promising approach to support patients with diabetes [1,2]. Telemedicine interventions in diabetes may include video consultations, text reminders, telemonitoring, etc. [3,4]. However, the effectiveness of diabetes-related telemedicine interventions in regards to patient-related outcomes needs evaluation.

The rapid development within telemedicine emphasizes the need for a new review. Moreover, previous systematic reviews describing the effectiveness of telemedicine in diabetes management focus on a specific type of telemedicine [5,6], a specific type of diabetes [7,8], specific comparators [2,6], or specific outcomes [2,9]. Thus, the present review has a broad scope with an eye to performing an exhaustive review within the field. The present systematic review and meta-analysis aims to evaluate the effectiveness of telemedicine interventions versus any comparator on diabetes-related outcomes among adult patients with diabetes.

2. Method

The present review considers studies that include adult subjects with a diagnosis of diabetes (type 1, 2, or gestational). Studies that evaluate telemedicine interventions are considered. These telemedicine interventions may include various technologies, and the remote communication may be managed by various health care professionals, trained peers, or be fully automatic. The review considers randomized controlled trials comparing a telemedicine intervention to any control that does not include telemedicine. The considered outcomes include all patient-diabetes related outcomes. Peer-reviewed full-text papers in English, Norwegian, Danish, and Swedish are considered.

¹ Corresponding Author; E-mail: svh@hst.aau.dk

A thorough search was performed in PubMed, CINAHL, EMBASE, and the Cochrane Library Central Register of Controlled Trials (CENTRAL). To provide an estimate of the treatment effect, the studies will be pooled via statistical meta-analysis.

3. Results

The search identified 11,759 potential papers. After removal of duplicates, 7,558 papers remained. These papers underwent screening (title/abstract), and 941 remained for full text reading. This resulted in a final sample of 218 papers. Preliminary results indicate that effects are limited in regards to type 1 diabetes and gestational diabetes. However, the effect in regards to type 2 diabetes shows more promise.

4. Discussion

The review is strengthened by the fact that the search of literature is very exhaustive. An expected limitation is the limited heterogeneity between telemedicine studies, as the studies differ in use of technology, organizational setup, etc., which complicates the comparison of studies.

5. Conclusion

The present review may add value to the existing evidence base by providing an exhaustive review of telemedicine interventions in diabetes. Further results will follow.

References

- [1] Kaufman N, Salahi A. Using Digital Health Technology to Prevent and Treat Diabetes. *Diabetes Technol Ther* 2017;19(S1):59–73.
- [2] Faruque LI, Wiebe N, Ehteshami-Afshar A, Liu Y, Dianati-Maleki N, Hemmelgarn BR, et al. Effect of telemedicine on glycated hemoglobin in diabetes: A systematic review and meta-analysis of randomized trials. *CMAJ* 2017;189(9):341–64.
- [3] Crico C, Renzi C, Graf N, Buyx A, Kondylakis H, Koumakis L, et al. mHealth and telemedicine apps : in search of a common regulation. *Ecancermedalscience* 2018;12:853.
- [4] Mahar JH, Rosencrance GJ, Rasmussen PA. Telemedicine: Past, present, and future. *Cleve Clin J Med* 2018;85(12):938–42.
- [5] Holtz B, Lauckner C. Diabetes Management via Mobile Phones: A Systematic Review. *Telemed e-Health* 2012;18(3):175–84.
- [6] Yun JE, Park JE, Park HY, Lee HY, Park DA. Comparative Effectiveness of Telemonitoring Versus Usual Care for Heart Failure: A Systematic Review and Meta-analysis. *J Card Fail* 2018;24(1):19–28.
- [7] Lee SWH, Ooi L, Lai YK. Telemedicine for the management of glycemic control and clinical outcomes of type 1 diabetes mellitus: A systematic review and meta-analysis of randomized controlled studies. *Front Pharmacol* 2017;8:1–8.
- [8] Kongstad MB, Valentiner LS, Ried-Larsen M, Walker KC, Juhl CB, Langberg H. Effectiveness of remote feedback on physical activity in persons with type 2 diabetes: A systematic review and meta-analysis of randomized controlled trials. *J Telemed Telecare* 2019;25(1):26–34.
- [9] Hu Y, Wen X, Wang F, Yang D, Liu S, Li P, et al. Effect of telemedicine intervention on hypoglycaemia in diabetes patients: A systematic review and meta-analysis of randomised controlled trials. *J Telemed Telecare* 2019;25(7):402–13.

The ‘Back Office’ of a Dispensing Cabinet: Technology and Work Contributing to Medication Safety

Valentina LICHTNER^{a,b,1}, Mirela PRGOMET^b, Bryony Dean FRANKLIN^{a,c},
and Johanna I WESTBROOK^b

^aDepartment of Practice and Policy, UCL School of Pharmacy, UK

^bCentre for Health Systems and Safety Research, AIHI,
Macquarie University, Australia

^cCentre for Medication Safety and Service Quality, Pharmacy Department,
Imperial College Healthcare NHS Trust, UK

Abstract. Automated dispensing cabinets in clinical wards may contribute to improving safety by reducing the likelihood of medications not being available when needed. However, achieving this safety benefit is dependent on a ‘back office’ sociotechnical infrastructure that combines semi-automated processes with mindful, resilient work practices.

Keywords. Patient Safety, Hospital Drug Distribution Systems.

1. Introduction

Missed medication doses (omissions) are a common medication administration error and negatively impact the safety of hospital inpatients (e.g. [1, 2]). One reason leading to omissions is unavailability of medication on the ward at the time of need [3].

Ward-based automated dispensing cabinets (ADCs) can assist with the management of ward medication stock (‘imprest’) and thus may reduce the number of doses omitted due to unavailability [4]. However, little is known about the ‘behind the scenes’ work that contributes to this effect. In this poster we describe and discuss the sociotechnical system that lies behind ward-based ADCs. This sociotechnical system usually remains invisible [5] to ADC users in clinical areas.

2. Methods

We draw on ethnographic research carried out in hospitals in the UK and in Australia. The UK study was conducted in an adult hospital in 2016-17, while the study in Australia is in progress. Both studies were wide-ranging, focused on clinicians’ medication work, with attention paid to medication activities carried out by pharmacy staff and technology such as ADCs. Activities were observed both in pharmacy areas

¹ Corresponding Author. Email: valentina.lichtner@mq.edu.au

and clinical wards. We interviewed pharmacists and technicians, as well as pharmacy managers and IT staff. In this analysis, we focus on pharmacy processes for management of imprest medications in ADCs. Our analysis was influenced by the *sensitivity to operations*, *preoccupation with failures* and *resourcefulness* that characterize resilient activities in high-reliability organizations [6], and by the literature on infrastructures in organizations [7].

3. Findings and discussion

ADCs are one of the many information technologies (IT) hospital pharmacies have in place to manage medication stock. Such IT include stock management applications, or systems to connect with manufacturers and wholesalers' catalogues. Some IT functionalities are automated: for example, the ADC alerts pharmacy staff of medications low in stock; a stock management system can be set up to automatically re-order routine stock from wholesalers.

Effective use of these systems, however, cannot rely exclusively on automation that works on data about past use and embedded rules. Instead, it requires pharmacy staff to pay attention to any emerging risks associated with the availability (or lack thereof) of different types of medication, foreseeing supply chain conditions, and resourcefulness in solving problems or constraints.

This is invisible, mindful [6] work of pharmacy staff behind the scenes of an ADC, that makes the replenishing of ADCs' imprest medication stock look seamless from the clinical areas, most of the time.

4. Conclusion

The 'backoffice' of hospital ADCs is a sociotechnical infrastructure that combines semi-automated processes with mindful, resilient work practices. It is this combination, rather than the technology on its own, that is likely to reduce the frequency of omitted doses for hospital inpatients.

References

- [1] C.N. Patel, M.D. Swartz, J.S. Tomasek, L.E. Vincent, W.E. Hallum, and J.B. Holcomb, The effects of missed doses of antibiotics on hospitalized patient outcomes, *J S R* **233** (2019), 276-283.
- [2] P. Rostami, C. Heal, A. Harrison, G. Parry, D.M. Ashcroft, and M.P. Tully, Prevalence, nature and risk factors for medication administration omissions in English NHS hospital inpatients: a retrospective multicentre study using Medication Safety Thermometer data, *BMJ Open* **9** (2019), e028170.
- [3] R.N. Keers, S.D. Williams, J. Cooke, and D.M. Ashcroft, Causes of Medication Administration Errors in Hospitals: a Systematic Review of Quantitative and Qualitative Evidence, *Drug safety* **36** (2013), 1045-1067.
- [4] C.M. Boyle, D.J. Maxwell, S.E. Meguid, D. Ouaida, and I. Krass, Missed Doses An Evaluation in Two Drug Distribution Systems, *The Australian Journal of Hospital Pharmacy* **28** (1998), 413-416.
- [5] S.L. Star and A. Strauss, Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work, *Comput. Supported Coop. Work* **8** (1999), 9-30.
- [6] K.E. Weick, K.M. Sutcliffe, and D. Obstfeld, Organizing for High Reliability: Processes of Collective Mindfulness, *Research in Organizational Behavior* (1999), 81-123.
- [7] S.L. Star, The Ethnography of Infrastructure, *American Behavioral Scientist* **43** (1999), 377-391.

The Prevalence of Alcohol and Tobacco Use Associated Risk Factor of Noncommunicable Diseases in Si Sa Ket Province, Thailand

Kampol KHEMTHONG^{a,1}, Niruwan TURNBULL^b, Savittri

Rattanopad SUWANLEE^c, Karl PELTZER^d

^{a,b}*Faculty of Public Health, Mahasarakham University, Thailand*

^c*Faculty of Humanities and Social sciences, Mahasarakham University, Thailand*

^d*Ton Duc Thang University, Ho Chi Minh City, Vietnam*

Abstract. Noncommunicable diseases (NCDs) were caused by risk factors also rising rapidly and killed more people. This study aimed to explore and determine the prevalence and distribution of alcohol and tobacco use as NCDs's risk factors. We used a cross-sectional survey on Health dataset between October 2013 and April 2017 of people who were 13 years old and older. This study was investigated included demographics, alcohol consumption and tobacco use. This study found risk of NCDs 49.07%. The majority of risk factor were men (50.2 %), age_group were 40-59 years old(24.4%), The most were men having diabetes mellitus family (43.0%), hypertension family (17.9 %), alcohol consumption (26.9%), tobacco use (19.0%), most of the women were high blood pressure (23.0%), high blood sugar level (33.3%), overweight and obesity (23.4%), waist was over (22.5%) and high total cholesterol (21.4%), alcohol consumption among the gender 37.8%, most were male (26.9%), age_group mostly 45-59 years old (19.3 %), married (23.1%), agricultural (29.7%), primary school (29.7%). The prevalence of risk factors, most risk factors was tobacco use in men (18.9%), OR 16.789, (95%CI, p-value<.001), alcohol consumption were men (26.9%), OR 3.934 (95%CI, p-value<.001).

Keywords. Health Data Centre, Health Dataset, Risk Noncommunicable diseases

1. Introduction

Noncommunicable diseases (NCDs) are incurable chronic diseases. People were death by NCDs caused a total of 41 million every year, up to 71% of all deaths globally, most death located in low and middle-income countries. The behavioural risk factors and metabolic factors are caused the major factors [1]. and underlying risk factors [1, 2]. Thailand had increased incidence of NCDs and death around 500,000 cases (71.0%) of all death[3]. Thai's hospitals collected electronic health data records and linkage with Health Database Centre (HDC) [4] This study aimed to explore and determined alcohol consumption and tobacco use in the prevalence associated with the risk factors of NCDs[5].

¹Corresponding author: E-mail: kampol.khe@msu.ac.th.

2. Methodology

We used a cross-sectional survey and data from electronic health data records were verified quality to ensured from all hospitals on Health Data Centre (HDC) database of Si Sa Ket Provincial Public Health. This studied data contained demographics, physical measurements, biological measurement. This dataset excluded all cases with missing, duplicate, error information and a case that a physician had diagnosis be NCDs patient. We selected population by purposive sampling who was 13 years old and older.

3. Results

Found risk people 49.07 %, were men (50.2 %), age_group were 40-59 years old(24.4%)(mean 50.97, S.D. 21.747), men having diabetes mellitus (43.0%), hypertension family (17.9 %), alcohol consumption (26.9%), tobacco use (19.0%), most of the women were high blood pressure (23.0%), high blood sugar level (33.3%), overweight and obesity(23.4%), waist was over (22.5%) and high total cholesterol (21.4%). alcohol consumption among the gender 37.8%, mostly were male (26.9%), female(10.9%), the group of 45-59 years old (19.3 %), married (23.1%), agricultural (29.7%), primary school (29.7%). The prevalence of risk factors, most risk factors was tobacco use in men (18.9%), OR 16.789, (95%CI 16.399-17.188, p-value<001), alcohol consumption were men (26.9%), OR 3.934 (95%CI 3.885-3.983, p-value<.001).

4. Discussion and Conclusion

A large electronic health data were studied from the Health Data Centre (HDC), it was a linkage of all Health server database permitted for administrators to verified data and sharing. We found a population that registered, but more cases were missing a health check data and electronic health data was an error, missing and duplicate. People who risk NCDs and the majority were men. Health evidence based on electronic health data could be support and provides health policy maker and effort to health personnel.

References

- [1] World Health Organization. Noncommunicable diseases 2019 [updated [Internet]. 2019 [cited 2019 May 18]. Available from: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>.
- [2] Baldock K, Paquet C, Howard N, Coffee N, Hugo G, Taylor A, et al. Associations between resident perceptions of the local residential environment and metabolic syndrome. *Journal of environmental and public health*. 2012;2012.
- [3] World Health Organization. World Health Statistics 2015 (http://www.who.int/gho/publications/world_health_statistics/2015/en). Accessed; 2015.
- [4] Ministry of Public Health. Health Data Center 2015 2015 [Available from: http://hdcservice.moph.go.th/hdc/main/index.php?sele_kpiyear=2015&level=1&sele_kpi=442f204be9d6aec9da15786ad707d5a4].
- [5] World Health Organization. WHO | About 9 voluntary global targets. WHO. 2018.

Toward a Harmonized WHO Family of International Classifications Content Model

Samson W. TU^{1a} Csongor I. NYULAS^a Tania TUDORACHE^a Mark A. MUSEN^a
Andrea MARTINUZZI^b Coen van GOOL^c Vincenzo della MEA^d Christopher G.
CHUTE^e Lucilla FRATTURA^f Nick HARDIKER^g Huib ten NAPEL^c Richard
MADDEN^h Ann-Helene ALMBORGⁱ Jeewani Anupama GINIGE^j Catherine SYKES^h
Can CEKIK^k Robert JAKOB^k

^aStanford University, USA, ^bIRCCS E. Medea Scientific Institute, Italy, ^cNational Institute for Public Health and the Environment, The Netherlands, ^dUniversity of Udine, Italy, ^eJohns Hopkins University, USA, ^fCentral Health Directorate, Friuli Venezia Giulia Region, Italy, ^gUniversity of Huddersfield, UK, ^hUniversity of Sydney, Australia, ⁱNational Board of Health and Welfare, Sweden, ^jWestern Sydney University, Australia, ^kWorld Health Organization

Abstract. An overarching WHO-FIC Content Model will allow uniform modeling of classifications in the WHO Family of International Classifications (WHO-FIC) and promote their joint use. We provide an initial conceptualization of such a model.

Keywords. ICD, ICF, ICHI, Family of International Classifications

1. Introduction

The 11th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-11) represents a fundamental change in the methodology for developing new classifications at the World Health Organization (WHO). Major departures from past ICD revisions include an information model of ICD categories called the Content Model, the maintenance of all-inclusive knowledge about ICD-11 content, called the Foundation, and the ability to create specialized classifications, called linearizations, for a range of use cases.

In addition to ICD, the WHO-FIC reference classifications include the International Classification of Functioning, Disability, and Health (ICF)[1], and the emerging International Classification of Health Interventions (ICHI)[2]. The goal of this work is to develop a generalized WHO-FIC Content Model that defines the information structure of the entities in the classifications and their semantic roles.

2. Method

The Content Model defines the types of entities and relationships in the WHO-FIC Foundation. We analyzed the structure of ICD, ICF, and ICHI to determine an

¹Corresponding Author, Samson W. Tu, Stanford University, USA; E-mail: swt@stanford.edu.

overarching conceptualization.

3. Results

We conceptualize the generalized WHO-FIC Content Model as consisting of the following: (1) *Codable entities* in a classification: These are the categories that represent phenomena that a classification is designed to encode. For ICD, diseases, disorders, and injuries are examples of codable entities. For ICHI, health interventions are the codable entities. For ICF, functioning situations (defined in terms of ICF domains and the impairment/limitation/restriction qualifiers) and environmental factor roles, will be codable entities. (2) *Structural component*: These are the entities that help to define the meaning of codable entities. For ICHI, targets, actions, and means are the structural components of health interventions. For ICF, body structure, body function, activities and participation, and environmental factors are the structural components that help to define the functioning situations and environmental factor roles that ICF is designed to code. No structural component is defined for ICD at this time. (3) *Defining qualifiers*: In ICF, some attributes, such as the extent of impairment qualifier, are necessary for the definition of codable entities. (4) *Extension codes*: These are entities that can be used to refine the codable entities. (5) *Properties of the classification entities*: These include informational properties, such as title and definition, and post-coordination properties that represent axes along which the entities can be defined logically or specialized.

In addition, a *post-coordination model* specifies how the structural components, defining qualifiers, and extension codes can be used, in conjunction with post-coordination properties, to define and refine codable entities. Finally, a *linearization model* specifies what entities and relationships are necessary to describe any linearization in the Foundation. It allows the specification of the entities that should be included in a linearization, the post-coordination axes are allowed or required in a given linearization, and constraints on the values of the post-coordination axes in a linearization.

4. Discussion

ICD and ICHI can easily be formulated in this framework. For ICF, the introduction of the concept of codable entities as composed from ICF domains and qualifiers formalizes what has been informally defined. The post-coordination model defines the semantic roles that each type of entity plays in the information model. In conclusion, the unification of WHO reference classifications in a common modeling framework is promising and provides a solid foundation for future enhancements of these classifications.

References

- [1] World Health Organization, International Classification of Functioning, Disability, and Health: ICF, Geneva, 2001.
- [2] World Health Organization, *International Classification of Health Interventions (Draft in Development)*, <https://www.who.int/classifications/ichi/en/>.

Towards an Ontology for Trustful mHealth Apps

Hanane DERBOUZ ROUIBETE and Stefano BONACINA^{a1}

^aHealth Informatics Centre (HIC), Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Abstract. A unified and integrated approach to represent mHealth apps and their characteristics is currently lacking. To fill this gap, the overall purpose of this project is to develop an ontology, to help address the objective of building 'trustful' mHealth apps. This paper is a brief presentation of the followed methods, and the preliminary results of the research, i.e. a first version of that ontology.

Keywords. mHealth apps, ontology development process, Unified Modeling Language Class Diagrams, Web Ontology Language, Protégé

1. Introduction

Despite the continuous growth of mobile health applications (mHealth apps), the available ones are problematic, due to the lack of a unified framework, specially devised for their development and quality assessment [1]. This prevents a common adoption of mHealth apps in healthcare delivery, as they might be risky. Massive accreditation programmes of mHealth apps seem to be unfeasible [2], though assessment frameworks have been developed [3] and development of standards is in progress (e.g. HL7 cMHAFF [4]). Alternative strategies for quality assessment have to be explored, as well. Biomedical Informatics research has demonstrated a successful development of ontologies. As defined in Computer Science, they are sets of concepts and their relationships, to represent a domain, by sharing structured, computable knowledge (e.g. the Gene Ontology [5], for genes' annotation). This project aimed to propose an ontology to help address the objective of building 'trustful' mHealth apps, sharing common understanding of the concepts, among people and applications.

2. Method and Results

We applied a combination of methods to develop an ontology for mHealth apps. The Design Science research methodology [6] has been involved as a main theory in the research process. We performed a literature search and analysis to identify features and characteristics of mHealth apps (search strings: "ontology development", "mHealth", "mHealth apps", "mobile apps", "content analysis", "methodologies for ontology development", sources: PubMed, Web of Science, and Google Scholar; time span:

¹ Corresponding Author,. Email: stefano.bonacina@ki.se.

2013-2017, language: English). The Unified Process for ONtology building (UPON) [7] and the Ontology Development 101 [8] guide have been applied for implementing the ontology. Protégé 5.2.0 and Astah UML have been used to represent the ontology by OWL 2.0 language and UML class diagrams, respectively. Finally, to identify subsumption relationships between classes, HermiT OWL Reasoner has been applied.

Sixteen papers were selected and analysed. The developed ontology includes 13 concepts and their relationships. Main concepts are “mHealth app” (with its features), “User”, “Development Process”, “Developer”, “Organisation”, “App Publisher”, and “App Documentation”, with sub-concepts. Main relationships are “use”, “assess”, “recommend”, “develop”, “is_supported_by”, “regulate”, and “provide”. OWL 2.0 format and UML class diagrams have been created, and consistency achieved.

3. Discussion and Conclusion

A first edition of an ontology for mHealth apps has been implemented, and its consistency achieved. Compared to past research [9], in this study mHealth apps and their development were the focuses, including OWL and UML format of the ontology.

This work has certain limitations; firstly, only two researchers performed the development. Secondly, literature from 2013 to 2017 was considered to identify concepts and relationships. Future work addresses involving more researchers and including updated literature.

4. Acknowledgments

This research has been supported by the Health Informatics Centre (HIC), LIME Department, Karolinska Institutet, Stockholm, Sweden.

References

- [1] C.K. Kao, D.M. Liebovitz, Consumer Mobile Health Apps: Current State, Barriers, and Future Directions, *PM R.* 9 (2017), S106-S115.
- [2] P. Wicks, E. Chiauzzi, 'Trust but verify'--five approaches to ensure safe medical apps, *BMC Med.* 13 (2015), 205.
- [3] R. Nouri, S. R. Niakan Kalhori, M. Ghazisaeedi, G. Marchand, M. Yasini, Criteria for assessing the quality of mHealth apps: a systematic review. *J Am Med Inform Assoc* 25 (2018), 1089-1098.
- [4] HL7 Consumer Mobile Health Application Functional Framework (cMHAFF), Release 1. [Online, June 2018] Available at: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=476
- [5] The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong, *Nucleic Acids Res.* 47(D1) (2019)D330-D338.
- [6] K. Peffers, T. Tuunanen, C.E. Gengles, et al., The design science research process: A model for producing and presenting information systems research, *First International Conference on Design Science Research in Information Systems and Technology*, 2006, 83-106.
- [7] A. De Nicola, M. Missikoff, R. Navigli, A Proposal for a Unified Process for Ontology Building: UPON, *Lecture Notes in Computer Science*, (2005)655–64.
- [8] N.F. Noy, D.L. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*, Stanford University, Stanford, CA, 2001.
- [9] J.D. Cameron, A. Ramaprasad, T. Syn, An ontology of and roadmap for mHealth research, *Int J Med Inform* 100 (2017),16-25.

Towards Prediction of Type 1 Diabetes Patients Who Fail to Achieve Glycemic Target

Morten Hasselstrøm JENSEN^{1a,b}, Simon CICHOSZ^a, Ole HEJLESEN^a, Irl.B. HIRSCH^c, Peter VESTERGAARD^{a,d}

^aSteno Diabetes Center North Denmark, Aalborg University Hospital, Denmark

^bDepartment of Health Science and Technology, Aalborg University, Denmark

^cDepartment of Medicine, University of Washington, Seattle WA

^dDepartment of Clinical Medicine, Aalborg University, Denmark

Abstract. In this study, we investigated which predictors from people with type 1 diabetes at initiation of intensive treatment that increase the risk of not achieving glycemic target. Data from a clinical trial with type 1 diabetes people ($n=460$) were used in a logistic regression model to analyze the effect of the predictors on achievement of glycemic target. Results indicate that age, smoking, glycosylated hemoglobin, 1,5-anhydroglucitol and fluctuation from continuous glucose monitoring are predictors of achievement of glycemic target, which can be used in an algorithm to predict people who fail to achieve glycemic target.

Keywords. Type 1 diabetes, prediction, continuous glucose monitoring

1. Introduction

Today's guidelines, such as, American Diabetes Association's Standards of Medical Care in Diabetes – 2019[1] recommend intensive treatment for diabetes patients to achieve a glycosylated hemoglobin (HbA_{1c}) level of less than 7% (54 mmol/mol). However, it is challenging for people with type 1 diabetes to achieve this glycemic target, and a too aggressive treatment increases the risk of severe hypoglycemia[2]. Prediction at initiation of intensive treatment of those who will fail to achieve target would enable an alternative personalized treatment with the potential of being more effective and safer. This study sought to investigate which predictors from people with type 1 diabetes at initiation of intensive treatment that increase the risk of not achieving glycemic target.

2. Method

Data from people with type 1 diabetes ($n=460$) enrolled in the Onset 5® trial by Novo Nordisk. Participants were on average 43 years old, 57% male, had a diabetes duration of 24 years, a body mass index of 26 kg/m² and an HbA_{1c} of 7.5% (58.5 mmol/L) at baseline. 359 (78%) of the participants did not achieve $HbA_{1c} < 7\%$ after 16 weeks of treatment. Time of randomization defined baseline and initiation of intensive insulin

¹ Corresponding author: e-mail: mhj@hst.aau.dk

treatment. A logistic regression model was constructed to analyze the effect of a variety of predictors at baseline on achievement of $\text{HbA}_{1c} < 7\%$ (Y/N) after 16 weeks of treatment. Predictors at baseline included age (stratified in 18-25, 25-35, 35-50, 50-65 and >65 years), gender, diabetes duration, body mass index, HbA_{1c} , smoking status, fluctuation from continuous glucose monitoring, 1,5-anhydroglycitol (1,5-AG) and fasting plasma glucose. Odds ratios (OR) are presented with 95% confidence intervals (CI).

3. Results

Higher age resulted in an increased risk of not achieving target (50-65 vs 18-25 years OR: 3.5, 95% CI: 1.3-9.6). Not surprisingly, higher HbA_{1c} at baseline increased the risk as well (OR: 14.7, 95% CI: 7.2-30.0). Being a smoker resulted in a 4-fold increased risk of not achieving target (OR: 4.0, 95% CI: 1.2-13.4). Increase in 1,5-AG resulted in a decreased risk (OR: 0.89, 95% CI: 0.87-1.01), whereas increased fluctuation in CGM increased the risk (OR: 1.34, 95% CI: 0.77-2.33), however, none of them were statistically significant. An analysis without fluctuation resulted in a statistically significant association between 1,5-AG and achieving glycemic target (OR: 0.87, 95% CI: 0.77-0.99).

4. Discussion

In this study, higher age affects ability to achieve glycemic target irrespective of diabetes duration. Furthermore, being smoker is associated with not achieving target, which is confirmed by Peng et al.[3]. Higher 1,5-AG level decreased the risk of not achieving target. Blood concentrations of 1,5-AG decreases during time of hyperglycemia, and a higher 1,5-AG level at baseline is thus a reflection of a better control with less glycemic variability. Furthermore, findings about fluctuation in CGM strengthen the indication of glycemic variability as a predictor of achievement of glycemic target.

5. Conclusion

Both age, smoking status, HbA_{1c} , 1,5-AG and CGM fluctuation are at baseline for people with type 1 diabetes who succeed or fail to achieve glycemic target after 16 weeks of treatment. These are valuable findings prior to development of an algorithm to predict people who do not achieve glycemic target from intensive insulin treatment where alternative treatment options should thus be considered.

References

- [1] American Diabetes Association, "6. Glycemic targets: Standards of medical care in diabetes - 2019," *Diabetes Care*, vol. 42, no. January, pp. S61-S70, 2019.
- [2] W. V. T. Rose A. Gubitosi-Klug, Barbara H. Braffett, Neil H. White, Robert S. Sherwin, F. John Service, John M. Lachin, "Risk of Severe Hypoglycemia in Type 1 Diabetes Over 30 Years of Follow-up in the DCCT/EDIC Study," *Diabetes Care*, vol. 40, pp. 1010-1016, 2017.
- [3] K. Peng et al., "Association between smoking and glycemic control in diabetic patients: Results from the Risk Evaluation of cAncers in Chinese diabeTic Individuals: A lONgitudinal (REACTION) study," *J. Diabetes*, vol. 10, no. 5, pp. 408-418, 2018.

Translating openEHR Models to FHIR

Georg FETTE^a, Max ERTL^a, Stefan STÖRK^a

^a*Comprehensive Heart Failure Center, University Hospital of Würzburg, Germany*

Abstract. openEHR and FHIR are two competing clinical data modeling and data exchange standards, that are commonly seen as mostly incompatible. However, the two have quite much in common and bridging approaches between the two worlds can serve the benefit of both communities. In the presented work, the data models of openEHR are translated into FHIR data models.

1. Introduction

FHIR and openEHR are two data model standards, which both have in common to use two-layer-modelling [1], i.e. they specify a meta-model, which is used to specify the domain model. If it were possible to translate openEHR models to FHIR models, many software components from the FHIR domain could be used to process openEHR data. Bosca [2] already described the possibility of using FHIR profiles as a reference model for openEHR and translating openEHR models to StructureDefinitions. The presented work analyzes to what extent an openEHR model can be generically and preferably without loss of information transformed into a FHIR model.

2. Methods

For the translation of an openEHR model all model parts of a given openEHR system (RM, archetypes, templates) are translated to StructureDefinitions. The set of StructureDefinitions representing the RM was created by hand. The ADL files of archetypes were parsed and all types inside fields were translated to individual elements of the translated StructureDefinition. The object-composition paradigm was realized by using BackboneElements. All elements of the anonymous class that represents a constrained archetypes element (as well as elements inherited from the RM type) were explicitly added to the BackboneElement. The identification of identifiable openEHR list elements was realized by incorporating the identifier into the path of the FHIR element. The cardinalities of elements were ignored and only the occurrences of the contained types were translated to FHIR. The translation of openEHR templates was realized by using the template format OET as translation source. The template files were parsed and for each contained archetype a FHIRPath constraint was added that required the existence of the element. All further constraints on the contained archetypes were also realized via FHIRPath constraints and added to the template model's root element. The described translation methodology was implemented and applied on the models on the international CKM.

3. Result

The translated archetypes and templates were successfully translated to StructureDefinitions and were validated without errors. By realizing the sub elements of archetypes as BackboneElements, the RM type of these elements was lost.

4. Discussion

The archetype translation was modeled using BackboneElements and by explicitly expanding the anonymous classes representing the archetype elements. The necessity to expand the anonymous classes make the logical models look quite bloated and complex when the used RM types themselves contain many elements (e.g. most RM types used in archetypes derive from LOCATABLE, which already adds 6 elements).

The template translation was modeled alternatively using FHIRPath constraints to realize object-composition. This solution has the drawback that type checking could become complicated, because the actual elements are not explicitly defined. Although FHIR validators are capable of checking these constraints, this option moves the workload of conformance checking to the FHIRPath part of the validator, which could probably reduce performance.

The OET format is commonly only used during template design time. It was developed for the Ocean Template Designer and is no official openEHR format. However, it has the advantage that all object-compositions and constraints on elements are still explicitly listed so they can be directly translated to FHIRPath constraints.

When using FHIR it is forbidden to create a new set of custom profiles. It creates systems that are semantically incompatible with other FHIR systems not using those models. The presented work aims purely on the one-to-one model translations and semantic mappings are not its scope.

5. Conclusion

An approach was presented to translate openEHR archetypes and templates to FHIR logical models. The approach was tested on the models contained in the international openEHR model repository CKM. The authors consider the translation of openEHR models to FHIR models a difficult issue, because the two standards use different object-composition, derivation and constraining paradigms. Either the resulting models contain a very large number of elements or the number of resulting models becomes very large.

References

- [1] Michelsen L, Pedersen SS, Tilma HB, Andersen SK. Comparing different approaches to two-level modelling of electronic health records. *Stud Health Technol Inform.* 2005;116:113-8.
- [2] Bosca D, Moner D, Maldonado JA, Robles M. Combining Archetypes with Fast Health Interoperability Resources in Future-proof Health Information Systems. *Stud Health Technol Inform.* 2015;210:180-4.

Uncovering Care Barriers Experienced by Patients and Loved Ones in the Emergency Department: Can a Health App Help?

Azzra MANGALJI^{a,1}, Kendall HO^{a,b,c}, Wayne CHOI^{a,b}, Chad KIM SING^{a,b}, Heather LINDSAY^{a,b}, Christopher Wai-Chung LEE^{a,b}, Kent LOFTSGARD^d, Michael LIM^c, and Nooshin JAFARI^c

^a University of British Columbia, Faculty of Medicine, Vancouver BC, Canada

^b University of British Columbia, Emergency Medicine, Vancouver BC, Canada

^c University of British Columbia, Digital Emergency Medicine, Vancouver BC, Canada

^d British Columbia Support Unit, Vancouver BC, Canada

Keywords. emergency department, care barriers, health technology, health app, patient satisfaction

1. Introduction

There are approximately 13.9 million ED visits in Canada each year. These visits can be overwhelming for patients and their loved ones. This study aimed to identify factors that diminish the ED experience for patients, guiding us to identify care barriers and informing us how to design and implement a person-centred app for patients visiting ED.

2. Methods

This mixed-methods study involved an online survey (n=125), focus groups (n=16) and individual interviews (n=14). Participants were patients and accompanying family or friends over 19 years who had visited any ED in BC in the past five years (June 2014 – June 2018). The survey, representing 169 visits to the ED, contained questions based on common care barriers identified in our pilot study[1] and past literature.

3. Results

Overall, 67.4% of interview patients were satisfied with the quality of care. However, we identified 5 ED care barrier themes presented below in order of prevalence of comments: **Overall Experience (34.1%)**: 44.3% felt communication challenges with ED staff, and 35.1% felt their care was unsatisfactory at one or more points during their visit.

Wait times (25.7%): Most (35.8%) spent 1-3 hours waiting to see a doctor, and 18.2% spent more than 10 hours total in the ED.

Discharge process (15.7%): 63.9% of the participants felt they were mentally and physically ready to manage their condition at home, whereas 36.1% were not.

¹ Corresponding author E mail: azzra.mangalji@alumni.ubc.ca.

Information Exchange (14.5%): Participants felt they did not receive enough information about amenities (74.6%), wait time (71.5%), wait reason (58.5%), who was going to be seeing them (56.6%), and aftercare instructions (52.4%).

Process Flow (10.0%): 64.7% of these participants were unclear as to next steps as they progressed through the ED, and 35.3% desired improvement in efficiency.

Use of Technology: When asked whether they felt an app would be helpful in their ED journeys, 47.5% of survey participants were sure they would use it, and 38.3% would decide depending on the functionality. Overall, 86.7% of interview participants wanted to use such an app if it existed, and those who were reluctant reported reasons such as security concerns and it being overwhelming to be a patient and expected to be on their phone.

4. Discussion

The care barriers identified were about timely communication, but were unrelated to the satisfactory quality of care experienced by the majority of patients. Suggested improvements that emerged were related to adjusting communication and administrative efforts, as follows:

Improving physician-patient communication: Interpersonal skills during communication ensures patients feel being the focus of care, aided with practical solutions such as physician name tags and context-appropriate touch

Improving patient experience of waiting: Electronic notifications to patients for their turn to see a physician untethers them from the waiting room. Furthermore, periodic checking to update patients of wait times or, “ED rounding,” can improve satisfaction, reduce frustrations, and decrease patients leaving prematurely.

Improving patient anticipation of ED journey: Providing a general road map to prepare patients for their journeys through the ED and updating them on what they are waiting for can help decrease logistical confusion and ease unrest.

Improving patient preparedness to manage their health post-discharge: Providing documentation of discharge summaries and home management information for patients to review at home can reduce anxiety, improve adherence to therapies and follow-up, and reduce unnecessary ED revisits.

It is important to note that ED overcrowding, experienced by most EDs, can compromise communication over delivery of quality care as health professionals have finite time. Therefore, solutions that support better communication and improve ED flow would be ideal.

5. Conclusion

We are now developing an ED app based on the priority care barriers identified, which are likely similar in other EDs. We aspire to have this app to be useful and scalable to other locations to optimize person-centred care.

References

- [1] Chen M (Ming YC.), Improving Patient Experience during Emergency Department (ED) Journey, 2016. Canadian Institute for Health Information, NACRS Emergency Department Visits and Length of Stay, 2018-2019, (2019).

Usability Issues with Mechanical Ventilation Devices

Beatrice COLDEWEY^a, Annette DIRUF^b, Rainer RÖHRIG^a
and Myriam LIPPRANDT^{a,1}

^a*Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Germany*

^b*Division for Medical Informatics, Carl von Ossietzky University of Oldenburg, Germany*

Abstract. The user interface of a mechanical ventilator is safety critical, as use errors can lead to patient harm. A systematic review was conducted to identify published usability issues and contributing factors that can lead to use errors. The findings were grouped in an Ishikawa diagram. Many of the problems mentioned based on inconsistent labeling and manufacturer-specific naming of ventilation modes. In the studies, usability was often measured quantitatively and did not allow any conclusions to be drawn about concrete problems.

Keywords. Ventilators, mechanical; usability problems; patient safety

1. Introduction

Due to technological progress in the medical field, more and more functionalities of mechanical ventilators lead to an increase in complexity, which is reflected in the user interface. To reduce use-associated risks, the application of usability engineering to medical devices is subject to legal regulations [1]. Despite use of approved equipment, use errors caused by perception, cognition or action error can cause adverse events with potential patient harm. The aim of this work is the systematic analysis of published usability issues and contributing factors that can lead to use errors described in literature.

2. Methods

A systematic review according to the PRISMA reporting standard was carried out. In PubMed a search string consisting of a combination and synonyms of the terms "mechanical ventilation", "usability" and "risk" was used to find eligible publications. Articles were included if detailed information on usability issues and contributing factors as cause for reported incidents or possible use errors with mechanical ventilation devices was given. The selected articles were qualitatively analysed by two independent researcher. The findings were systematically listed in an Ishikawa diagram by the occurring main categories.

¹ Corresponding Author.; mlipprandt@ukaachen.de

3. Results

3.1. Study Selection

The search strategy returned 379 articles. The titles and abstracts were assessed based on the inclusion and exclusion criteria leaving 31 studies for the full text screening. Two papers could not be obtained as full text. After reading the remaining articles, 19 papers were omitted due to exclusion criteria. There are 10 articles left after this procedure.

3.2. Use Error Analysis

The 10 articles eligible for inclusion deal with the usability evaluation of ventilators and approaches for alternative forms of presentation as well as the analysis of critical events. The main causes for use errors were inconsistent labelling of the ventilation settings and parameters across different devices (e.g. manufacturer / language specific terminology), aspects of the design (e.g. the position of the power switch, the design of interaction elements like buttons or the complexity of and divergent interaction models cross different devices) and the absence of experience and training according to manufacturer specification with the ventilators [2–5].

4. Discussion und Conclusion

This review showed that there are several usability issues related to mechanical ventilation systems described and discussed in the current literature. The terms usability and user error were not used uniformly. Most of the articles have dealt with the usage of the devices, but not with the causes of the use errors. Many of the found usability problems were caused by inconsistent labelling across the devices and different manufacturers [2–4]. Therefore, the implementation of an internationally standardized nomenclature for ventilation modes and parameters is recommended [2–4]. Most of the studies focused on quantitative methods and statistical evaluations. Potential causes were given only as examples as explanation for the quantitative results in the discussion. For the analysis of use errors new testing methods and a systematic causal analysis should be performed. Further investigation on usability problems of mechanical ventilators and contributing factors is recommended.

References

- [1] IEC 62366-1:2015-02, Medical devices - Part 1: Application of usability engineering to medical devices.
- [2] M. Jiang, S. Liu, J. Gao, Q. Feng, and Q. Zhang, Comprehensive Evaluation of User Interface for Ventilators Based on Respiratory Therapists' Performance, Workload, and User Experience. *Med Sci Monit* **24** (2018), 9090–9101.
- [3] N. Marjanovic and E. L'Her, A Comprehensive Approach for the Ergonomic Evaluation of 13 Emergency and Transport Ventilators. *Respir Care* **61** (2016), 632–639.
- [4] J. Spaeth, T. Schweizer, A. Schmutz, H. Buerkle, and S. Schumann, Comparative usability of modern anaesthesia ventilators: a human factors study. *Br J Anaesth* **119** (2017), 1000–1008.
- [5] F. Templier, P. Miroux, F. Dolveck, A. Descatha, N.-S. Goddet, C. Jeleff, M. Baer, M. Chauvin, L. Brochard, and D. Fletcher, Evaluation of the ventilator-user interface of 2 new advanced compact transport ventilators. *Respir Care* **52** (2007), 1701–1709.

User Satisfaction with a Speech-Enabled Translator in Emergency Settings

Antony Akash JANAKIRAM^a, Johanna GERLACH^b, Alyssa VUADENS-LEHMANN^a, Pierrette BOUILLON^b and Hervé SPECHBACH^a

^a Geneva University Hospitals, Department of Community Medicine, Geneva, Switzerland

^b Faculty of translation and interpreting (FTI), University of Geneva, Geneva, Switzerland

Abstract. In medical emergency situations, the language barrier is often a problem for healthcare quality. To face this situation, we developed BabelDr, an innovative and reliable fixed phrase speech-enabled translator specialised for medical language. Majority of participants (>85%) showed a positive satisfaction level using BabelDr.

Keywords. Fixed phrase speech-enabled translator, web application, user satisfaction, emergency settings

1. Introduction

Following the refugee crisis, European hospitals are increasingly confronted with patients who have no language in common with the medical staff. At the Geneva University Hospitals (HUG), 52% of the patients are foreigners and 8% speak no French at all. Miscommunication is a known threat to quality, security and equitability of medical care [1, 5]. Professional interpreters are not always available in emergency settings and considered to be expensive [2]. Generic technologies such as Google Translate (GT) are not reliable enough [3]. Medical fixed phrase translator lack of usability. In this context, the HUG in collaboration with the FTI, supported by the Private foundation of HUG, developed BabelDr, a speech-enabled fixed-phrase translator. Like other systems, it relies on pre-translated sentences, but includes speech recognition, allowing doctors to freely ask questions instead of searching for them in a list. Pretest showed that BabelDr is significantly more precise than GT and presents higher usability than MediBabble [4]. The aim of the present study is to evaluate BabelDr in real emergency settings. We focus on two aspects: global user satisfaction and the perceived quality of patient-doctor communication through the system, from the patient's and doctor's point of view.

2. Method

Patients were recruited between January and August 2019 in the outpatient emergency unit of the HUG. Inclusion criteria are: patients must speak a language available on BabelDr and must not have a French level sufficient for a standard diagnostic interview.

Patients are recruited in the outpatient emergency unit of the HUG. Exclusion criteria are patients who aren't able to read in their language. Each doctor followed a short training before using BabelDr. Participants are enrolled during daytime, fill a consent form and receive instructions on how to interact with the tool. During the consultation, the use of the tool can be interrupted by the doctor or the patient. Doctors can use the speech recognition system or directly select sentences in a list to ask their questions and the patients are required to answer non-verbally. After the consultation, all participants (doctors and patients) fill in a satisfaction survey paragraph.

3. Results

In total, 22 patients (age: $M=41.18$, $SD=17.44$) are eligible for this study and speak: Spanish (9), Arabic (4), Farsi (6), Tigrinya (2) and Albanian (1). Results on the satisfaction level show predominantly a positive feedback from both patients and doctors (>85%). Regarding patient-doctor communication, 68.2% of patients think the system allows them to completely explain the reason for consulting and 90.1% of doctors think BabelDr allows them to understand the patient's problem.

4. Discussion

We observe that most patients are able to communicate their reason to visit and doctors mostly understand the problem, suggesting that BabelDr is suitable for communication in emergency settings. User satisfaction is higher for patients than doctors, who can feel constrained by the available questions. Patients' dissatisfaction is related to misunderstood translations (due to level of literacy, text-to-speech quality and dialects).

5. Conclusion

Based on the remarks of patients and doctors, we continue to develop the tool by expanding coverage and adding more languages. A bidirectional system, allowing patients to answer by selecting pictograms, is also in development.

References

- [1] G. Flores, M.B. Laws, S.J. Mayo, B. Zuckerman, M. Abreu, L. Medina, & E.J. Hardt, Errors in medical interpretation and their potential clinical consequences in pediatric encounters, *Pediatrics* **111**(1) (2003), 6-14.
- [2] G. Major, J. Napier, L. Ferrara, & T. Johnston, Exploring lexical gaps in Australian Sign Language for the purpose of health communication, *Communication & Medicine* **9**(1) (2012), 37-47.
- [3] S. Patil, & P. Davies, Use of Google Translate in medical communication: evaluation of accuracy, *Bmj* **349**, g7392
- [4] H. Spechbach, J. Gerlach, S.M. Karker, N. Tsourakis, C. Combescure, & P. Bouillon, A Speech-Enabled Fixed-Phrase Translator for Emergency Settings: Crossover Study. *JMIR medical informatics* **7**(2) (2019), e13167
- [5] M. Wasserman, M.R. Renfrew, A.R. Green, L. Lopez, A. Tan-McGrory, C. Brach, & J.R. Betancourt, Identifying and preventing medical errors in patients with limited English proficiency: key findings and tools for the field. *Journal for Healthcare Quality* **36**(3) (2014), 5-16.

User-Centered Design for Promoting Patient Engagement in Chronic Diseases Management: The Development of CONCERTO+

Marie-Pierre GAGNON^{a,b,1}, Mame-Awa NDIAYE ^a Alain LAROUCHE ^c, Guylaine CHABOT^c, Christian CHABOT^d, Ronald BUYL^e, Jean-Paul FORTIN^a, Anik GIGUÈRE^f, Annie LEBLANC^g, France LÉGARÉ^{a, f}, Aude MOTULSKY^g, Claude SICOTTE^g, Holly O.

WITTEMAN^f, Éric KAVANAGH^h, Frédéric LÉPINAY^h, Jacynthe ROBERGE^h, Hina HAKIM^f, Myriam BRUNET-GAUTHIER^h, Carole DÉLÉTROZⁱ, Samira A. RAHIMI^j, Jack TCHUENTE ^a, Maxime SASSEVILLE^{aj}.

^a *Centre de Recherche sur les Soins et les Services de Première Ligne, Québec, Canada*

^b *Faculty of Nursing, Université Laval, Québec, Canada*

^c *Groupe Santé Concerto, Québec, Canada*

^d *Patient expert, Québec, Canada*

^e *Department of Public Health, Vrije Universiteit Brussel, Jette, Belgium*

^f *Department of Family and Emergency Medicine, Université Laval, Québec, Canada*

^g *School of Public Health, Université de Montréal, Montréal, Canada*

^h *École de design, Université Laval, Québec, Canada*

ⁱ *School of Health Sciences (HESAV), HES-SO, Lausanne, Switzerland*

^j *Department of Health Sciences, Université du Québec à Chicoutimi, Canada*

Abstract. Multimorbidity increases care needs among people with chronic diseases. In order to support communication between patients, their informal caregivers and their healthcare teams, we developed CONCERTO+, a patient portal for chronic disease management in primary care. A user-centered design comprising 3 iterations with patients and informal caregivers was performed. Clinicians were also invited to provide feedback on the feasibility of the solution. Several improvements were brought to CONCERTO+, and it is now ready to be implemented in real-life setting.

Keywords. Patient portal, Chronic disease management, User-centered design

1. Introduction

Chronic diseases are the leading cause of mortality in the world, and account for nearly 70% of deaths [1]. Interventions increasing patient involvement in their own care are associated with better health outcomes [2]. eHealth technologies have the potential to support chronic disease management by considering end-users' needs, preferences, and capabilities. CONCERTO+ is an innovative solution to improve patient experience and includes personalized follow-up, health education, appointment management, and communication between patients, informal caregivers and primary healthcare teams.

2. Methods

The development of CONCERTO+ was done through a collaborative approach between patients, caregivers, researchers, clinicians, designers, and IT developers who identified the functionalities to include. The solution was developed through 3 phases, based on user-centered design methods [3], and usability was assessed with patients.

3. Results

- Phase 1: Development of the 1st prototype by a design team based on a review of available solutions. Usability test performed with 10 patients and informal caregivers using think-aloud and eye-tracking methods.
- Phase 2: Design of the 2nd prototype, addition of graphics and more functions. Usability tested with 9 people with a lower education level. Generally positive comments on CONCERTO+, and some improvements suggested.
- Phase 3: 3rd prototype designed in response to the results of the 2nd usability test. A 3rd usability test will be done to make the final adjustments to the solution.

4. Discussion

Users generally found CONCERTO+ intuitive and easy to navigate. After 2 iterations, only minor changes were brought to the design, although improvements are still possible in the final version. Chronic patients and their informal caregivers are willing to use CONCERTO+ to communicate with their primary healthcare team.

5. Conclusion

The final prototype will be translated into a fully interactive application with the support of our IT partners. A pilot trial will assess the feasibility of a large-scale study.

References

- [1] World Health Organization. WHO - Global status report on noncommunicable diseases. 2014. http://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf.
- [2] J. Greene, J.H. Hibbard. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. *J Gen Intern Med*. 2012; **27**:520-6.
- [3] K.R. Devi, A.M. Sen, K. Hemachandran. A working framework for the user-centered design approach and a survey of the available methods. *Int J Sci Res Publ* 2012;**2**(4):1-8.

Using Long Short-Term Memory (LSTM) Neural Networks to Predict Emergency Department Wait Time

Nok CHENG^{a,1} and Alex KUO^a

^a School of Health Information Science, University of Victoria, Canada

Abstract. Emergency Department (ED) overcrowding is a major global healthcare issue. In this paper, we used Long Short-Term Memory (LSTM) recurrent neural networks to build a model to predict ED wait time in the next 2 hours using a randomly generated patient timestamp dataset of a typical patient hospital journey. Compared with Linear Regression model, the average mean absolute error for the LSTM model is decreased by 15% (3 minutes) ($p < 0.001$). The LSTM model statistically outperforms the LR model, however, both models could be practically useful in ED wait time prediction.

Keywords. ED Overcrowding, ED Wait Time Prediction, Machine Learning, Linear Regression, Recurrent Neural Network, Long Short-Term Memory

1. Introduction

Emergency Department (ED) overcrowding is a complex, systemic and global healthcare issue resulting in decreased care efficiency and worsening patient outcomes [1]. ED wait time, defined as the time to physician initial assessment, can support decision making to enhance patient experience and improve hospital resource allocation [1]. Very limited research studies explore the use of *Long Short-Term Memory (LSTM) Recurrent Neural Networks* [2] in Emergency Medicine research, while LSTM has been commonly used in fields such as Business consistently providing more accurate sequence predictions [3]. Comparatively, *Linear Regression* (LR) is a popular and simple prediction method, but such a method may not be accurate in the complex ED overcrowding problem. The aim of this paper is to preliminarily investigate the application of the LSTM method in ED overcrowding analysis by comparing the prediction accuracy of ED wait time in the next 2 hours of an LSTM model and that of an LR model.

2. Methods

A dataset was created to represent 50,000 patient instances of timestamp differences randomly generated over a pre-defined range uniformly. A sequence of five (5) timestamp differences were used to illustrate the six (6) timestamps in a patient journey,

¹ Corresponding Author, Nok Cheng, School of Health Information Science, University of Victoria, HSD Building, A202, Victoria, B.C., Canada; E-mail: nokcheng@uvic.ca.

including: patient arrival time, triage time, registered time, physician initial assessment time, ED exit time, and inpatient unit discharge time. The average ED wait time for the next 2 hours was calculated in minutes. Two models were built: 1) a single-layered LR model with an input array of 1×5 for the 5 timestamp differences; and 2) a single-layered LSTM model consisting of 5 input units corresponding to the 5 timestamp differences in the sequence with the final unit outputting to a linear regression layer. Both models were trained and validated with 80% of the data using 100-Fold Cross Validation and tested with the remaining 20% of the data. Both models include an early stopping to prevent overtraining and use mean absolute error (MAE) as loss function. A two-sample t-test was performed on the MAE values from the testing phase with a significant difference of $p < 0.001$ and 99% confidence level.

3. Results

The LSTM model achieved a 15% (3 minutes) lower average MAE ($p < 0.001$), 7.49×10^{-2} (99% CI $5.57 \times 10^{-2} - 9.40 \times 10^{-2}$) (17 minutes) compared to 8.84×10^{-2} (99% CI $7.27 \times 10^{-2} - 10.41 \times 10^{-2}$) (20 minutes) in the LR model.

4. Discussion

Supported by Tax et al. 2017 [13], the LSTM outperforms the LR due to its ability to learn the relationships between each consecutive time-step. The 3-minute prediction difference between the 2 models suggests that either of the 2 models could be practically useful in wait time prediction. An ED wait time prediction model is a beneficial clinical decision support tool to provide patients control in their decision making of ED care [4], and improves care efficiency with proactive allocation of resources such as increasing workload nurses and overflow units [1]. A major limitation of the study is the use of randomly generated dataset without real patient flow and data quality representation.

5. Conclusion

Though the LSTM model statistically outperforms the LR model, both models could be practically useful in wait time prediction. Next steps include using broader patient data variables from a level 1 trauma hospital to explore different LSTM and/or deep neural network architectures and compare with various accuracy evaluations.

References

- [1] A. Affleck, P. Parks, A. Drummond, B.H. Rowe, and H.J. Ovens, Emergency department overcrowding and access block, *Canadian Journal of Emergency Medicine* **15** (2013), 359-370.
- [2] S. Hochreiter et al., Long short-term memory, *Neural computation* **9** (1997): 1735-1780.
- [3] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, Predictive Business Process Monitoring with LSTM Neural Networks, *Lecture Notes in Computer Science (LNCS)* **10253** (2017), 477- 492.
- [4] M.R. Friesen, R.D. McLeod, T. Strome, and S.N. Mukhi, Load Balancing at Emergency Departments using 'crowdinforming', In *Proceedings of the 13th Internal Conference on e-Health Networking, Applications and Services*, IEEE, (2011), 364-370.

Subject Index

-omics biomarkers	1163	aspirin	1383
A-Test method	178	assistive robots	1323
access control	663	assistive technologies	577
accessibility	1299	attitude	1098
accuracy	242	augmented reality	322, 453
acronym resolution	292	authoring tool	517
active electrodes	751	autism spectrum disorder(s)	287,
adherence	761		1399, 1401
administrative claims	1133	automatic data processing	133
adoption	1001	automation	158, 302
adverse drug events	198, 618, 683	automatization	926
adverse drug reaction(s)	48, 648	autonomous systems	1089
adverse events	163	Axon evaluation	1209
affective computing	509	BabelDr	527
aged	1081	ballistocardiography	479, 808
agent-based model (ABM)	1193	bar code medication administration	
aging	577, 678, 1168	(BCMA)	1279
air pollutants	1343	Bayesian modelling	422
airflow obstruction	731	bed management	377
alert	128	bed occupancy	1363
alert fatigue	906	behaviour change	1041, 1291
alert override	906	behaviour monitoring and	
algebraic Petri nets	597	interpretation	746
algorithms	228	bibliometrics	1118
Alzheimer cohorts	1183	big data	1123, 1163, 1307
Alzheimer's disease	499, 1108	big data analytics	489
ambient assisted living	746, 828	big data science	1381
analytic-decision modelling	198	bio-statistics	1215
anamnesis	43, 342	bioinformatics	1195
anatomical teaching	322	biomarkers	1195
android	3	biomedical data harmonisation	387
annotation guideline	28	biomedical genomics	267
anonymization	68, 193, 1138	blended learning	1325
antibiotics	63	blockchain	1293
app design	951	brain imaging	1168
application	1389	brainstem	1309
appropriateness	906	breast cancer	1243
apps	1041	breast cancer surgery	628
archetypes	228	Brief International Cognitive	
Arden Syntax	607	Assessment for MS	
artificial intelligence	113, 153, 307,	(BICAMS)	168
	382, 889, 1001, 1089, 1123,	Brief Visuospatial Memory Test-	
	1303, 1399	Revised (BVMT-R)	168

business intelligence	494	clinical laboratory	437
business models	1046	clinical models	567
business process		clinical narratives	427
management	522, 1275	clinical NLP	427
camera calibration	1245	clinical notes	587, 1355
cancer	941, 1223, 1255	clinical outcomes	542
cancer care	572	clinical practice guidelines	517, 628, 1313
capacity building	1311	clinical process	1193
carbohydrates	693	clinical research	78
cardio-mediastinal features	1225	clinical specialties	292
cardiomegaly	1225	clinical terminologies	18, 327
care barriers	1417	clinical text	28, 43, 148, 342
care coordination	1177	clinical trial	183, 252, 337, 961, 1098
care management	1006	clinical trial as topic	367
case identification	1369	clinical trial protocols as topic	367
case study	1036, 1385	clinical trial recruitment support	
case-based learning	484	system	158
case-based reasoning	537	clinical validation	494
CEN 13606 archetype	552	clinical workflow	906
certification	1094	clinicians	956
change management	971, 1265	clinico-environmental monitoring	103
chatbot evaluation	976	cloud computing	412
chatbot(s)	796, 976, 1361, 1399	cluster analysis	88
checklist	668	co-morbidity	422
chemotherapy	1217, 1271	co-production	623
child	1016	cognitive computing	1201
children's health	1297	cognitive impairment	1213
chronic disease management	1423	cohort catalogue	1183
chronic fatigue syndrome	98	cohorts	108
chronic obstructive pulmonary		collaborative	663
disease (COPD)	731	collaterals	33
chronic wound	607	color histogram analysis	1303
citizen reference	1168	colorectal cancer (CRC)	1191
citizens	956	combination prevention	848
claim data	853	Common Terminology Service	
claims auditing	879	Release 2 (CTS2)	1371
classification	257, 297, 362, 1227	communication	1056, 1323
clinical	981	community participation	1098
clinical abbreviations	88	community pharmacy services	803
clinical concept disambiguation	93	competencies	1143, 1311
clinical data integration	78	competency assessment	1311
clinical data management	1199	competency-based learning	1311
clinical decision support		complex and massive data	108
system(s) (CDSS)	73, 198, 517, 532, 542, 607, 623, 628, 648, 673, 683, 688, 1319	computable phenotypes	228
clinical informatics	98, 1325	computer graphics	1329
clinical information extraction	93	computer interpretable	
clinical interpretable guidelines	623	guidelines	517, 628
		computer vision	1245

computerized physician order entry (CPOE)	1217	data quality	118, 143, 183, 337
computerized provider order entry	1271	data reuse	218, 683
confidentiality	317, 1170	data safety	597
constraint solvers	688	data science	252, 489
consumer health informatics	1026, 1253	data security	597, 1170
consumption	1383	data sharing	392, 1133, 1161, 1235, 1393
content heterogeneity	387	data transfer model	557
contextual problem list	567	data visualisation	108, 218, 1239, 1397
contextual word embeddings	432	data warehousing	272, 547, 1185
continuity of care	223, 582, 633	de-identification	148
continuous EEG	1317	decision aids	1319
continuous glucose monitoring	277, 1413	decision making	1353
contrast agent	123	decision quality	562
controlled	1335	decision quality instruments	562
ContSys standard	223	decision support system(s)	158, 372, 618, 916, 981, 1001, 1291, 1227, 1303
conversational agent	1399	deep contextualized representation	88
conversational user interface	976	deep learning	73, 148, 153, 382, 823, 1223, 1241
convolutional neural network	168, 1241, 1355	delta check	494
cooperative sensors	751	dementia	746
coping	1229	dengue	853
critical care	73, 1273	depression	921, 1285
cross-mapping	1233	depressive symptoms	1285
CT angiography	123	design	718, 1213
cyber-physical systems	693	detrended moving average	1309
cytidine deaminase	1205	diabetes	693, 951, 1403
data abstraction	257	diabetes mellitus	352
data analysis	233, 317	diabetes self-management education and support	1263
data collection	367, 961	diabetic foot syndrome	1395
data concerning health	1138	diabetic kidney disease	1289
data discovery	317	diagnosis	731
data encryption	262	diagnosis-related groups (DRG)	1353
data exchange	557, 1163	diagnostics	1163
data governance	1401	dialysis patient risk factor	53
data integration	183, 262, 337, 1235, 1327	diet records	1016
data interoperability and data integration	267	digital biomarker	1163
data management	397, 653	digital consent	1103
data mapping	38	digital divide	1123
data mining	347, 402	digital ethics	1094
data privacy	1170	digital health	448, 567, 577, 869
data processing	183, 337	digital health intervention	708
data protection	68, 193	digital hospital	3
data provenance	1237	Digital Imaging and Communications in Medicine (DICOM)	663

digital mediation	1051	eLearning	1265
digital pathology	663	electrocardiogram	128
digital personalized health	828, 1365	electrocardiography (ECG)	198, 751
digital phenotyping	736	electrodermal activity	1269
digital slides	1223	electronic consent	1066
digital therapeutics	736	electronic data submission	23
digital tools for supplementing		electronic general consent	1151
medical teaching	322	electronic health record(s)	
digital visit	643	(EHR)	38, 43, 83, 98, 143, 148,
digitalization	638		163, 228, 247, 327, 382, 422, 469,
digitalize	168		484, 499, 567, 582, 587, 703, 818,
disability	1299		823, 901, 981, 1031, 1181, 1211,
disaster management	1321		1233, 1241, 1243, 1265, 1281,
discharge summaries	1341		1285, 1287, 1333, 1337
discrete choice experiment	1213	electronic health records/	
disease risk prediction model	1289	utilization	698
disease trajectory	469	electronic medical records	
distributed ledger technology	1293	(EMR)	342, 648, 718, 1261,
DNA methylation	1168		1345, 1383, 1397
docker	1393	emergency medical service (EMS)	613
doctor-patient relationships	1345	electronic medication administration	
document classification	48	system	1267
document retrieval	208	electronic patient records	1103
documentation	587, 771, 1281	electronic patient reported outcome	
downtime	1229	(ePRO)	941
drug combinations	63	electronic phenotyping algorithm	838
drug database(s)	63, 133, 1329	electronic prescribing	803, 833
drug misuse	133	electronic prescription	1351
drug-condition interactions	648	electronic whiteboard	188
drug-drug interactions	648	emergency	128, 1373
drug-gene interactions	648	emergency care	1001
dry electrodes	751	emergency communication	613
dyslalia	357	emergency department	1177, 1417
e-accessibility	453	emergency department	
e-glasses	453	overcrowding	1425
e-health	522, 552	emergency department wait time	
e-learning	322	prediction	1425
early mild cognitive impairment	1309	emergency documentation	613
economic analysis	961	emergency medical service	613
eConsent	3	emergency settings	1421
eConsult	1377	emotion	1269
education	781, 1143, 1325	EN13606	228
effects realization	971	epic	703
eHealth	3, 766, 776, 803, 818, 833,	ePRO and PROs functionality	1051
	848, 936, 941, 956, 1011, 1021,	equality	956
	1056, 1123, 1349, 1377	ergonomics	1076
eHealth literacy	1026	eSource	367, 961
elderly	1187, 1349, 1359	ethical issues	1189
elderly care	1046	ethical principles	1089

ethics	1098, 1189	genomic models	443
European Reference Network LUNG (ERN-LUNG)	986	genomic privacy	238
evaluation	542, 976, 1036	genomic sequencing	1061
evaluation study	1351	geographic mapping	1353
event	1275	geolocation tracking	103
evidence-informed		geometrical-coefficients	458
decision-making	858	geriatrics	542
exercise duration	658	German discharge summaries	28
experimental games	1215	glaucoma	153
expert advisory board	986	glaucomatous optic neuropathy	153
expert system(s)	618, 1207	glioblastoma	1207
experts reasoning	1313	global health	1098
exposome	1331	Google search	1333
external validation	838	Grading of Recommendations	
extract-transform-load (ETL)		Assessment, Development and	
tools	138	Evaluation (GRADE)	532
extreme value analysis	494	graft rejection	272
eye tracking	287	granularity	257
failure	1229	graph-based database	1253
FAIR data	392, 1237	gravitation	1353
family of international		Great East Japan Earthquake	1321
classifications	1409	ground truth	282
Fast Healthcare Interoperability		guidelines	532
Resources (FHIR)	3, 13, 138,	hand gesture	756
158, 474, 597, 618, 961, 991,		hand recovery	756
1185, 1395		health analytics	1231
fasting time	971	health and social care services	223
feature extraction	247	health app	1417
feature selection	257	health application	976
federated computation	1170	health care	13, 638, 1046, 1143,
first aid	1297		1307, 1385
fixed phrase speech-enabled		health care data	8
translator	1421	health care fraud	879
focus groups	941	health data	143, 402, 1237,
follow-up	916		1293, 1327
formative construct	562	health data centre	1407
framework(s)	906, 1076	health data science	233, 996
fraud detection	1307	health data sharing	894
free and open source software	981	health data warehouse	108
French	427	health dataset	1407
fundus image	153	health disparities	1123
gender	1181	health emergencies	1311, 1387
general data protection		health equity	1123
regulation (GDPR)	1046, 1118,	health facility	1301
	1293	health informatics	38
genome-wide association study		health information	1187
(GWAS)	412	health information	
		exchange	818, 1011

health information		hospital management	1363
interoperability	133, 367, 592, 1185	hospital service suspension	407
health information management	352	hospital-at-home	522
health information systems	463, 1036, 1229	hospitals	602
health information technology	633	human factors	1076
health insurance companies		human immunodeficiency viruses (HIV)	838, 848
(HIC) data	262	human-computer interface	1076
health issue network	484	human-robot collaboration	1245
health level seven (HL7)	818	human-robot interaction (HRI)	1323
health literacy	1026	HUMIRA	874
health monitoring	103	hypoglycemia	693
health policy analysis	833	I2B2	78
health professionals	966	iconic composition	791
health promotion	1187	identification	864, 1181
health services	577	image processing	272
health services delivery	678	imaging biomarkers	33
health smart home	828	impact on care	1323
health surveys	894	implementation	703, 971, 1279
health technology	113, 1417	inclusive design	926
health-enabling and ambient assisted technologies	746	indicators	1315
health-related data	1170	infection surveillance	1295
hearing loss	1031	inference	238
heart failure	761, 808	inference attacks	193
heart sound signal	178	informatics	1143
hepatitis	848	information	1108
hidden Markov model	287	information entropy	357
hierarchy	257	information extraction	203, 272
Higuchi's fractal analysis	1309	information governance	1401
history	1113	information model	1243, 1401
HIV prevention	1365	information storage and retrieval	63, 347, 813, 1381
HIV-indicator reporting	1301	information system	1177
HMN Tool	602	information system quality	602
home automation technology	746	information systems	613
home care	1359	information technology	1094, 1143
home care provider	723	informed consent	1138
home nursing	723	inpatient	188
home rehabilitation	1349	Instagram	781
homomorphic encryption	317, 412	instrument	587
Hosmer-Lemeshow test	504	insurance claims data	407
hospital admission	1359	integrated approach	848
hospital data	213	integrated care	223, 623, 633
hospital drug distribution systems	1405	integration	1315
hospital hygiene	1295	intelligent phonocardiography	178
hospital information systems (HIS)	557	inter-observer variability	736
		inter-rater agreement	282
		interconnected	1163
		interdisciplinarity	1113

International Classification of Diseases (ICD)	18, 78, 297, 417, 427, 1355, 1401, 1409
International Classification of Health Interventions (ICHI)	1409
International Consortium for Health Outcomes Measurement (ICHOM)	1243
international patient summary	552
internet	781
internet of things	693
interoperability	18, 402, 448, 474, 828, 926, 1163, 1170, 1315, 1395
interpretability	307, 1006
interpretations	916
interview	1021
ischemia	33
ISO 13606	1243
k-nearest neighbors	537
Kanta services	833
kidney disease	781
kidney transplant	272
knowledge resources	1387
knowledge to action	1387
knowledge transfer	1387
knowledge-based systems	352
laboratory test	916
large-scale interventions	833
latent Dirichlet allocation	422
lean	1036, 1071
leap motion	756
learning and development	1311
legislation	1046
legislation and jurisprudence	1128
length of stay	377
Levenshtein	1337
linear regression	1425
linguistics	347
linkage to care	848
literature	884
literature review	741, 1201
location based social networks	996
Logical Observation Identifiers Names and Codes (LOINC)	78, 138
logistic regression	163
long QT	198
long short-term memory	1425
long term monitoring	751
lower respiratory tract infection	1333
lung cancer	1319
machine learning	73, 168, 198, 242, 282, 302, 312, 327, 372, 382, 489, 499, 509, 761, 864, 889, 1163, 1221, 1225, 1227, 1357, 1391, 1425
machine learning method	1191
management information systems	1353
master patient index	1337
mechanical	1419
medical and healthcare informatics	1113
medical artificial intelligence	242
medical coding	312
medical data transformation	377
medical diagnostics	537
medical education	484
medical image	23
medical informatics	108, 347, 547, 869
medical informatics education	981
medical information retrieval	208
medical records	342, 1247
medical schools	1215
medical wearables	766
medication	1251, 1355
medication administration	1267
medication effects	874
medication error	1036
medication information	1026
medication management	926
melanoma	1303
mental health	1391
mental illnesses	746
metadata	1237
metric learning	1006
mHealth	332, 776, 796, 848, 894, 926, 946, 1016, 1273, 1361, 1365
mHealth apps	1411
micro-activity	673
microbiological data	397
microsimulation	233
midwifery	1239
mining social media	874
missing data	387
misspelling correction	43
mixed reality	322
mobile application(s)	698, 889, 991, 1016, 1066, 1297, 1373

mobile health	786	nursing minimum data sets	38
mobility	996	nursing technology	1305
models	1353	nutrition assessment	1016
models of care	678	obesity	864
mortality	1273	obstructive sleep apnea	1381
mosquito	332	occupational health care	1281
motivation	894, 1098	occupational health nursing	1281
movement disorders	889	older adults	1213
multi-criteria decision support	532	oncology	307, 1161, 1217, 1271
multi-lead	751	online	1108
multicenter clinical study	23	online consultation	643
multilayer perceptron	1219	online learning	1387
multilingual matching	18	online symptom checkers	966
multimorbidity	469, 623, 688	ontologies harmonization	1371
multiple disease progression		ontology(ies)	417, 517, 791, 858, 1327
model	469	ontology alignment	13
multiple sclerosis	168, 658, 911	ontology development process	1411
muscle fatigue	1219	ontology mapping	297
mutations	443	ontology-based information	
national data repositories	833	extraction	879
national health programs	818	ontology-based systems	352
national survey	1056	open standards	158
native XML database	552	openEHR	443, 592
natural language processing		OpenLink Virtuoso	417
(NLP)	43, 53, 83, 93, 98, 163, 203, 208, 272, 292, 302, 342, 347, 402, 432, 843, 911, 1195, 1211, 1341, 1381	OpenWHO	1387
natural language		operational confidentiality of	
understanding	432, 976	information	8
needs assessment	1081	operative report	382
negation detection	342	operative surgical procedure	668
neonatal screening	597	ophthalmological diseases	1197
NetLogo	1193	organization and administration	1128
neural network(s)	83, 756, 1207, 1209	osteoporosis	532
neuro-developmental disorder	1401	out-of-hospital cardiac arrest	1357
neuroimaging	1221	outbreaks	1387
neurointensive care unit	1317	outlier detection	1307
neurosurgery	163, 382	overweight	864
next generation sequencing		OWL	417
(NGS)	397	ownership	1301
non-verbal communication	791, 1329	P5 medicine	1089
nosocomial infections	1295	picture archiving and communication	
nurse role	786	systems (PACS)	663
nurse(s)	1021, 1267, 1279	pain measurement	946
nursing	188, 1323	palliative care	941
nursing documentation	1233	paper health record	1345
nursing homes	542	paramedics	613
nursing informatics	1233, 1275, 1305	paraplegia	741
		participatory design	941
		participatory management	1305

patient access to electronic health		phenotype computation	1327
records	698, 1021, 1056	phenotyping	499
patient behaviour	996	phone reminder	1251
patient data management system		physical activity	1287
(PDMS)	1295	PM10	1343
patient empowerment	781, 1061, 1361	policy	1108, 1133
patient engagement	1066	polynomial trendline	357
patient experience	1071	polypharmacy therapies	648, 688
patient harm	771	population synthesis	233
patient involvement	1056	portable health clinic	1347
patient journey	1071	postpartum depression	1391
patient participation	698	precision medicine	813, 823, 884
patient path	218	prediction	257, 731, 1359, 1413
patient portal(s)	1031, 1423	prediction model	307, 489
patient preferences	931	preference-sensitive	562
patient privacy	8	preferences learning	1313
patient readmission/statistics and		prenatal care	582
numerical data	547	prevention	848, 1197, 1297
patient safety	118, 668, 1405, 1419	primary health care	643, 1197, 1359, 1369
patient satisfaction	1417	privacy	317, 448, 926, 1011, 1128, 1161
patient search identification		privacy by design	653
service	1337	privacy-conscious tools	238
patient sitters	673	private-cloud	1170
patient-centered	1123	probabilistic data integration	387
patient-generated health		problem list	567
data	766, 936, 946	problem oriented medical record	567
patient-reported outcome		process mining	522
(PRO)	1051, 1255	process modelling	1193, 1275
patient-reported outcome		professional competencies	1199
measures	991	professional development	1231
perceptions	188	prognostic accuracy of score	504
performance	1301	prognostic models	1273
perinatal indicators	213	prognostic score	504
personal data	1138	prostate cancer	1241, 1369
personal health care	123	protected health information	1293
personal health library	448	protein diversity	1205
personal health record(s)	103, 698, 931, 1011, 1347, 1375	protein modelling	1331
personal health report	1373	prototype	901
personal medical information	1373	protégé	417, 1411
personalization	509, 926, 1006, 1385	pseudonymization	262
personalized health (pHealth)	1089, 1151, 1156, 1161, 1163, 1168, 1170	psoriasis	1375
personalized medicine	648, 1061	psycho-social factors	1041
Petri net(s)	484, 607	psychosocial	1399
pharmacogenetics	618	PU Learning	864
pharmacogenomics	267, 618	public health	796, 843, 894
pharmacovigilance	48, 198, 1227	public health care	1397
phase space networks	1219	public health informatics	796

public health professional		return of spontaneous circulation	
education	1215	(ROSC)	1357
public health surveillance	853	rheumatoid arthritis	532
public interest	1138	risk management	1128
PubMed	292	risk noncommunicable diseases	1407
qualitative research	633, 668	risk quantification	238
quality assurance	143	RNA-editing	1205
quality control	58	robotics	1245
quality improvement	1071	rosacea	1381
quality management	638	safety	1267, 1367
quality of care	542	safety management	771, 1247
quality of clinical notes	587	Sankey diagram	218
quality of data	267	scalability	193
R software	494, 1397	schizophrenia	1221
radiology report(s)	58, 203	scientific IT services	1170
random forest	1309, 1357	scientometrics	1118
randomized control trial	1399	score ISS-RTS-TRISS	504
range of motion	1389	second primary cancer (SPC)	1191
rare disease register	552	secondary use of electronic health	
rare diseases	986	records (EHT)	38, 78, 158, 1138
rare respiratory diseases	986	secure broker	332
real-world data	53	secure multi-party computation	8
real-world evidence	1369	security	412, 1161
receiver-oriented encryption		seismocardiography	479, 808
(ROE)	653	self-help groups	1375
recommendation systems	1183	self-management	1041, 1291
reconstruction	458	semantic interoperability	78, 223,
record	1383		437, 552
recurrent neural network	382, 1425	semantic similarity	13
refilling	1251	semantics	463, 1335
relation extraction	874	sense inventory	292
relational similarity	874	sensitive data	1170
reliable computing	68	sensor(s)	128, 673, 741
renal	458	sentence alignment	362
repolarization disorders	198	sentiment analysis	911
reproducible	1393	sepsis	1163, 1195
research	1098	septal heart defects	178
research data	1237	serious games	509, 736, 1215
research data management	392	serious mental illness	1367
research infrastructure	1170	service design	1385
ResearchStack	3	service virtualization	1170
resource allocation	1363	sexually transmitted infections	
resource description framework		(STIs)	848
(RDF)	417	shared decision making	1319
respiratory disease	1343	shelter	1321
responsible innovation	1189	sickle-cell disease	1361
Responsible Research and Innovation		sieve-based model	93
(RRI)	1189	simulation	1193
restorative dentistry	1201	single nucleotide variant	1331

skill	1143	symptom management	1255
skin cancer	1303	syndromic surveillance	843
smart home	746	systematic reviews	302
smartphone	1389	systems architecture	352
smartphones apps	1367	Tacotron 2	527
SNOMED CT	78, 138, 818, 1243, 1395	technology acceptance	1061
social anxiety	713	technology transfer	113
social care	1143	telehealth	776, 828, 1377
social determinants of health	173, 474	telemedicine	572, 761, 776, 828, 1377, 1403
social fear	713	telemonitoring	1403
social functioning	1399	telerehabilitation	658
social media	786, 921, 1081, 1391	teletriage	776
social participation	1081	terminology{ies}	173, 437, 552, 884, 921, 1233
social phobia	713	terminology as topic	1335
social robots	1046	terminology mapping	1335
social support	1261	testing	848
Society for Clinical Data Management (SCDM)	1199	text analysis	58
software design	463, 931	text data mining	48, 53, 312, 761, 884, 1253
software validation	592	text simplification	362
soundex	1337	text-to-speech	527
spellchecker	43	theoretical models	1076
spiking deep belief network	1269	therapy	1323
spinal cord injury	741	time growing neural network	178
standardization	1235, 1243, 1347, 1395	timeline	247
standardized language	1275	tomography	33
standardized nursing terminology	38	topic modelling	422, 911
statistical data interpretation	1215	Torsades-de-Pointes	198
stem cells	1235	toxicogenomics	1331
strategic planning	869	traditional medicine	791
stroke	509	training program	1231
STROKE OWL	653	transgender	901, 1181
structured clinical knowledge	638	translational bioinformatics	1331
structured data	443	translational informatics	474
subsequence time-series		trend analysis	252
clustering	277	triage advice	966
supervised machine		TrustZone	332
learning	48, 547, 1247	tuberculosis (TB)	848
support vector machines	1221	tumour burden	628
surveys and questionnaires	786, 921, 956	Twitter	911
survival analysis	377	two-channel biosensor	1317
Swiss Personalized Health Network (SPHN)	1151, 1156, 1161, 1163, 1168, 1170	Type 1 diabetes	1413
SWOT	869	unified medical language system	1335
symptom assessment	936	unified modeling language class diagrams	1411
		universal HIV testing	838
		usability	1076, 1275

usability evaluation	453	visual impairment	453
usability inspection	1066	visualization	247, 718
usability problems	1419	vital signs	1211
use cases	1046	vocabularies	173
user experience	1076	vocabulary	1335, 1393
user interaction	372	wavelet transform	128
user requirements	703	wearable device	1317
user research	951	wearable sensors	751
user satisfaction	1421	wearables	332, 1367
user-centered design (UCD)	901, 926, 1103, 1423	web accessibility	1108
user-computer interface	1081	web application	397, 1421
validation	242	web application for home care	723
validation study	213	web ontology language	1411
validity of prognostic scores	504	website	1299
value-based healthcare	572, 708	whole-slide imaging (WSI)	663
variant	884	wiki word vectors	148
variations	443	women's health	1391
ventilators	1419	word embedding(s)	43, 93, 208
video consultation	643	work environment	1021
virtual coach	509	workflow	1036
virtual reality	713	workstation on wheels	1279
		young men	1365

Author Index

Abadie, D.A.	818, 1011	Antons, R.	746
Abanos, S.	552	Antunes, R.	93
Abdellatif, A.	542	Argibay, O.	1011
Abouelenein, S.	1199	Arief, M.	1108
Abou-Khalil, V.	718	Arizpe-Gomez, P.	1245
Abu-Hanna, A.	257	Arkhipov, I.	33
Adams, K.	1076	Armagnague Thivant, C.	1345
Ademoye, K.	567	Årsand, E.	693, 796, 894, 1041
Aeschimann, J.	397	Artemova, S.	108
Aguirre Ojea, F.	1211	Artola, G.	517
Ahmed, A.	1347	Arvanitis, T.N.	623
Ahonen, O.	1143	Asikainen, P.	1267
Ainsworth, J.	1103	Askari, M.	781
Aizpurua, M.	1231	Atique, S.	153
Akhlaghpour, S.	1118	Attias, M.	1311, 1387
Akinyemi, O.	1339	Auchter, S.	607
Albert, F.	1353	Avillach, P.	18
Alegre, F.	1011, 1265	Ayala, F.	818, 1011
Alegria, J.	103	Babic, A.	178, 1301, 1323
Aleman, M.	1243	Bailey, J.E.	448
Alemu Gelaye, K.	1251	Bajekal, M.	469
Alexander, D.C.	469, 499	Balabaeva, K.	43, 342
Alexander, L.T.	1170	Balcaen, T.	247
Alexander, N.	499	Baldner, J.	1199
Ali, S.M.	946	Balestra, G.	1193
Allouch, S.B.	1046	Bamdeg, J.	1229
Almborg, A.-H.	1409	Bangerter, A.	287
Almeida, J.R.	93, 1183	Bankava, A.	1138
Alonso, E.	78	Bannay, A.	547
Alpay, L.	1349	Baqué, M.I.	818
Alsarkhi, A.	1203	Barbié, V.	884
Álvarez, R.	1011	Barbier, E.	108
Alzate, C.	173	Barda, J.	1170
Amantea, I.A.	522	Barkhausen, J.	123
Ammar, N.	448	Barkhof, F.	499
Ammenwerth, E.	38, 463	Barmina, T.	33
Ammour, N.	367	Barone, A.	1201
Amorese, M.	1337	Barro, S.G.	602
Amr, A.	1061	Barzdins, J.	1138
Anastasiou, A.	143, 509	Batey, D.S.	1365
Andersen, N.S.	951	Bathelt, F.	597
Andreata, D.	1233	Battegay, M.	1163
Ankolekar, A.	1319	Bauer, J.G.	613

Baum, A.	582	Bo, M.	522
Baum, A.J.	1345	Bochud, M.	1168
Baumberger, D.	38	Bodin, M.	108
Bavendiek, U.	808	Boella, G.	522
Bazile, P.	1353	Bolliger, C.	1170
Beach, M.	623	Bonacina, S.	1411
Bedouch, P.	108	Böning, A.	1295
Bedrick, S.	813	Borbolla, D.	474
Belhadj, I.	417	Borchiellini, A.	1193
Belkind, U.	1365	Borgwardt, K.	1163
Bellafqira, R.	412	Born, J.	1297
Belmin, J.	542	Borstnik, U.	1170
Ben Kimoun, I.	1227	Borycki, E.M.	1076
Ben Said, M.	1401	Bosson, J.-L.	108, 1215
Bender, M.	613	Bota, V.	723
Benitez, S.	188, 587	Bott, O.J.	272
Bennie, M.	1327	Botto, A.	587, 901
Benning, N.-H.	741	Bouaud, J.	542, 623, 628, 698
Benzenine, E.	213	Boudemaghe, T.	417
Berian, D.	357	Bouillon, P.	527, 1421
Berlanga, A.	1319	Bousquet, C.	267, 1227
Berlin, M.	1377	Bousquet, T.	1377
Bermudez, A.	1377	Bouzillé, G.	18, 427
Bernard, G.R.	874	Bouzille, G.	247, 547
Bertelsen, P.	956, 1051	Bowles, J.	688
Berthelot, H.	1329	Bowrin, P.	1263, 1399
Bettencourt-Silva, J.H.	173	Bradway, M.	1041
Beukenhorst, A.L.	946	Brat, H.	58
Beuscart, J.-B.	247, 683	Bräunlich, G.	1170
Bianca, D.	522	Brenck, F.	1295
Bienzeisler, J.	1128	Bricault, I.	108
Bignens, S.	168	Brinkmann, A.	1245
Bild, R.	68, 193	Brisimi, T.S.	879
Billey, K.	1335	Brix, T.J.	592
Bindel, M.	463	Brochhausen, M.	1089
Birchmeier, M.E.	168	Brody, R.	1066
Bird, J.	327	Broekhuizen, Y.	1257
Bischof, A.	123	Brou, K.M.	791
Bjelogrljic, M.	198, 208	Bruce, I.	387
Blach, C.	337	Bruce, J.	1365
Black, A.	1311	Bruchanski, L.	931, 1375
Black, C.	1327	Bruggemann, L.	1331
Blagosklonova, E.	33	Brunet-Gauthier, M.	1423
Blancafort, C.	628	Brunetti, E.	522
Blanes-Selva, V.	864	Bruyneel, A.-V.	1389
Blaser, J.	3	Büchler, A.C.	1163
Blasini, R.	158	Buckeridge, D.L.	858
Blatier, J.-F.	108	Budrionis, A.	148
Blobel, B.	352, 1089	Bühlmann, P.	1163

Buka, A.	1138	Chazard, E.	18, 218, 247, 547, 683
Bukowski, R.	1391	Chéhelat, O.	751
Burdisso, N.	838	Chen, C.-H.	1006
Bürgin, R.	38	Chen, H.	53, 252
Burgun, A.	432	Chen, R.	623
Bürkle, T.	397, 494, 618, 673	Chen, T.	874
Bussien, C.	1016	Chen, W.	1327
Butt, Z.	53	Chen, Y.-C.	1191
Buyl, R.	803, 1423	Cheng, N.	1425
Bygholm, A.	951, 1051	Cherqaoui, Z.	1401
Byrde, O.	1170	Chirila, O.S.	756
Cabitza, F.	242, 282	Choi, W.	1417
Cajander, Å.	1021	Choi, Y.	736
Calafiore, M.	247	Chomutare, T.	148
Calandra, T.	1163	Chute, C.G.	1409
Calvier, F.-É.	267	Cichosz, S.	1413
Campagner, A.	242, 282	Cichosz, S.L.	1179
Campos, F.	818	Cilar, L.	1273
Cancé, C.	108	Cirkovic, A.	1325
Cantora, A.	1265	Cisse, F.	1401
Caporossi, A.	108	Ciulu, M.D.	1303
Cardon, R.	362	Claerhout, B.	78
Cardone, M.B.	1341	Claes, M.	911
Carevic, L.	1325	Claveau, V.	427
Cariou, M.	367	Coatrieux, G.	412
Carter, D.	843	Cohard, B.	108
Caskey, F.J.	936	Coldewey, B.	1419
Cassarino, M.	901, 1181, 1279	Colussi, G.	587, 668, 931, 1375
Castiglione, R.	1011	Coman Schmid, D.	1170
Castillo, C.	921	Consorti, F.	484
Cattin, P.	1168	Corbett, A.	1076
Caucheteur, D.	884	Cornelisse, L.	1081
Cavin, D.	1161	Cornet, R.	981
Cejas, C.A.	818, 1011	Correa, E.	901, 1181
Cekik, C.	1409	Cort Madsen, I.	v, vii
Cha, W.C.	1373	Cosmaro, L.	848
Chabot, C.	1423	Costa, C.	103, 663
Chabot, G.	1423	Cottenet, J.	213
Chachere, D.	1203	Couch, P.	387
Chalder, T.	98	Courtine, M.	63
Chang, C.-C.	1191	Cozzi, E.	1377
Chanoine, S.	108	Crameri, K.	1170
Charlet, J.	1335	Crepain, T.	78
Charriere, K.	108	Crisan-Vida, M.	128
Chaste, P.	1401	Crooks, C.J.	422
Chatellier, G.	367	Cruz, J.	1243
Chatterjee, A.	1259	Cruz-Correia, R.J.	1315
Chatterjee, M.	287	Cuendet, M.	1161
Chauvet, R.	1227	Cuggia, M.	18, 427, 547

Cui, W.	489	Despotou, G.	623
Cullen, C.	879	Dessie Gashu, K.	1251
Curcin, V.	228	Deuber, D.	1061
da Silva Junior, G.B.	781	Dhalluin, T.	547
Daley, B.J.	1239	Dhrangadhariya, A.	58, 302
Dalianis, H.	148	Di Marco, L.	1215
Dalleur, O.	683	Diaz Maffini, M.	1211, 1337, 1377
Dalloux, C.	427	Dicko, A.	602
Damachi, F.	208	Dimitriadis, V.K.	848
Daniel, C.	367	Diruf, A.	1419
Danilov, G.	163, 382	Dixon, W.G.	946
Darmoni, S.J.	1335	Djadi-Prat, J.	367
Das, S.	1006	Doherty, P.	387
Dasen, S.	751	Dolotova, D.	33
Dauvillier, J.	1170	Donitova, V.	33
David-Tchouda, S.	1215	Donovan, K.M.	337
Davis, R.L.	448	Dormosh, N.	257
De Blas, A.	623	Dornauer, V.	463
De Bruijn, B.	843	Dorronzoro Zubiete, E.	911
De Freitas, S.	377	Dourado, D.X.C.	781
De Lame, P.-A.	602	Dowie, J.	532, 562
de Lusignan, S.	1339, 1369	Draganski, B.	1168
De Manuel, E.	623	Drake, J.	1391
de Matos Lima, S.	1217, 1271	Drenkhahn, C.	437
de Oliveira, J.G.R.	781	Dube, K.	1239
de Ruysscher, D.	1319	Düblin, P.	1151
De Schepper, K.	78	Dufossez, F.	1353
de Vasconcelos Filho, J.E.	781	Dugas, D.	322
Decaudin, B.	683	Dugas, M.	592, 889
Decorsaire, L.	1016	Dunstan, J.	347
Degen, G.-A.	494	Duong, V.	1151
Dekker, A.	1319	Dutta, R.	98
Del Fiol, G.	337	Dziekiewicz, M.	458
Del Zotto, M.	1098	Eberl, I.	38
Délétroz, C.	1423	Eckert, P.	1163
Delgado-Gonzalo, R.	332	Egelseer-Bründl, T.	761
Della Mea, V.	297, 1223	Egger, C.	1061
Della Ventura, M.	357	Eggerth, A.	761
Demuro, P.	1123	Egli, A.	1163
Denaxas, S.	469, 499	Ego, A.	1215
Denecke, K.	613, 796, 976	Ehrler, F.	1016, 1081
Deng, R.	1205	Eicher, J.	193
Denis, J.-L.	1001	Eils, R.	443
Derbouz Rouibete, H.	1411	Eisenmann, U.	347
Derchi, G.	1201	Eisenstein, E.L.	961
Dermitzakis, M.	1168	El Amrani, M.	218
Descalzo, J.	1279	Elkin, P.	823
Descloux, C.	1016	Elkin, P.L.	1205, 1381
Deserno, T.M.	1237	Ellingsen, G.	703

El-Zein, Y.	238	Frattura, L.	1409
Engelsma, T.	1213	Freundler, N.	1161
Engler, R.	1170	Frexia, F.	443
Ercia, A.	936	Frey, U.	1156
Eriksen, J.	1051	Frid, A.S.	776
Ertl, M.	1415	Froelicher, D.	317
Escobar Lopez, P.	1170	Frund, J.	337
Esteban, S.	582, 838, 1231, 1261, 1283, 1285, 1287, 1333, 1341, 1345, 1383, 1397	Fuad, A.	853
Eterno, J.	108	Fuge, I.	1305
Ezat Wan Puteh, S.	906	Fuglerud, K.S.	926
Faithfull, S.	1369	Fujarski, M.	889
Falls, Z.	1205	Fukuyama, M.	638
Farrington, K.	936	Funkner, A.	43, 342
Fatehi, F.	713, 1118	Furlong, L.I.	921
Fauconnier, J.	108	Furrer, H.	1163
Favre, M.	1329	Fürschke, A.	123
Faxvaag, A.	1293	Furuya, H.	638
Fechner, M.	592	Gabarron, E.	796, 911
Fellay, J.	238, 1161	Gadde, M.	118
Feng, S.	874	Gagnon, M.-P.	1423
Fenton, N.E.	1239	Gall, H.	138
Fernandez-Llatas, C.	522	Gallos, P.	1197, 1307, 1351
Ferrareis, L.	1011	Galluzzi, T.	1231, 1397
Ferrario, D.	751	Galvan, C.	1279
Ferrero, F.C.	1333	Gamhewage, G.	1387
Ferretti, G.	108	Gamhewage, G.M.	1311
Fette, G.	1415	Gan, H.N.	1357
Ficheur, G.	247, 683	Ganapathy, N.	1269
Fick, P.	1319	Ganzhorn, S.	1365
Fifelski, C.	1245	Ganzinger, M.	347
Fijacko, N.	1273	Garcelon, N.	432
Fijten, R.	1319	Garcia, G.	668
Finkelstein, J.	489, 658, 1066, 1235	García, M.T.	367
Fischer, H.	1128	Garcia, N.	1243
Fischer, P.	138	García-Gómez, J.M.	864
Flade, I.	1061	Garmann-Johnsen, N.F.	453
Flegel, V.	1170	Garofalo, R.	1365
Fleitmann, M.	123	Garrison, K.A.	1041
Fortin, J.-P.	1423	Garvey, A.	367
Fotteler, M.	1207	Garza, M.Y.	337, 961
Foufi, V.	48, 1098	Gassino, F.	818, 1011
Fournier, D.	58	Gaudet-Blavignac, C.	48, 322, 1098
Fournier, K.	843	Gavrilov, A.	33
Fox, M.	1170	Gebel, S.	1371
Fraccaro, P.	1367	Geifman, N.	387
Franco, M.	582	Geleijnse, G.	307
Franklin, B.D.	1405	Gencer, B.	198
		Genin, E.	412
		Gentili, N.	367

Georgakopoulos, S.	1307	Grippo, F.	297
Georgakopoulou, S.	1170	Grosjean, J.	1335
George, R.	1387	Grossmann, R.	3
Gerdes, M.W.	1259	Gruhl, M.	1393
Gerlach, J.	123, 527, 1421	Guessant, S.	133
Germer, S.	1185	Guevel, C.	818
Gesicho, M.B.	1301	Guhres, M.	643
Gfeller, S.	673	Gulden, C.	158
Ghalandari, M.	463	Günther, V.	494
Gharehbaghi, A.	178	Gütschleg, H.	479, 808
Giacomelli, G.	443	Güttler, K.	607
Giacomini, M.	1371	Habli, I.	1367
Giguère, A.	1423	Hachey, P.	843
Gillois, P.	1215	Hackl, W.O.	38
Gimenez, C.	1271	Haenni, E.	751
Gimenez, C.L.	1217	Hafsa, I.	1401
Ginige, J.A.	1409	Hagen, N.	741
Giordano, N.	1193	Häggglund, M.	1031
Giraldo, L.	188	Hahn, T.	889
Girardin, F.	198	Hahn, U.	28
Girlea, A.	723	Haida, K.	277, 1289
Giudicessi, R.	587	Hakim, H.	1423
Giunti, G.	911, 1385	Halminen, O.	113
Glachs, D.	1291	Handels, H.	123
Gligoroski, D.	1293	Hangaard, S.	731, 1403
Gobeill, J.	312, 884	Hänninen-Ervasti, R.	966
Goda, K.	407	Hansske, A.	1353
Godtliebsen, F.	148	Hardiker, N.	1409
Goldberg, S.	372	Hardt, T.	592
Goldman, J.-P.	48, 208	Hartvigsen, G.	693, 894, 1361
Gomel, L.	587	Hassan, F.A.	98
Gómez De La Camara, A.	367	Hassandoust, F.	1118
Gómez, A.	78	Hasselgren, A.	1293
Gonçalves, L.S.	1233	Hatanaka, J.	1363
Gonzalez Alvaredo, E.	1011	Hattori, A.	23
Gonzalez Pannia, P.	1333	Haugaard, L.H.	951
González, L.	78	Haux, R.	746
Gonzalez, N.	623	Hawthorne, C.	1195, 1331
Gordon, G.S.	961	Hayn, D.	761
Gosak, L.	1273	Hechtel, N.	1305
Goueslard, K.	213	Heikkinen, J.	1170
Govindarajan, S.	1225	Hein, A.	233, 653, 1245
Grabar, N.	362, 427	Hejlesen, O.	731, 1179, 1403, 1413
Grande Ratti, M.F.	901, 1181	Hejlesen, O.K.	1359
Grande, M.	668	Helfer, S.	597
Greub, G.	1163	Hellrich, J.	28
Griebel, L.	1061	Helou, S.	718
Griessbach, A.	3	Helsing, E.	1031
Griffon, N.	367	Hendriks, M.P.	307

Henriksen, A.	894	Ishankulov, T.	163
Heponiemi, T.	966	Ishii, T.	1321
Hernández, G.	78	Islam, M.	153, 1241
Hernández, L.	1243	Islam, R.	1347
Hersh, W.R.	813	Isleyen, F.	557
Hertzum, M.	703, 971	Íslind, A.S.	996
Heuser, F.	297	Issom, D.	1098
Hidalgo, M.A.	1365	Issom, D.-Z.	894, 1361
Hilbey, J.	367	Izukura, R.	1347
Hilfiker, R.	302	Jacques, L.	1016
Hinderer, M.	1061	Jacquot, M.	1170
Hiraga, C.	1187	Jaeger-Schmidt, C.	443
Hiragi, S.	1247, 1363	Jafari, N.	1417
Hirose, M.	1247	Jagannatha, S.	287
Hirsch, Irl.B.	1413	Jahn, F.	463
Hirshfield, S.	1365	Jähne-Raden, N.	479, 808
Hirvonen, P.	708	Jais, J.-P.	1401
Hitman, G.A.	1239	Jakob, R.	1409
Ho, A.F.W.	1357	Jakob, S.M.	1163
Ho, K.	1417	James, B.	946
Höffner, K.	463	Janakiram, A.A.	1421
Holban, S.	1303	Jansen, T.	307
Holl, F.	1207	Janson, A.L.	926
Hollis, K.F.	813	Jaspers, M.	1257, 1277
Holm, J.	397, 494, 673	Jaspers, M.W.M.	1213
Homco, J.	1076	Jauer, M.-L.	1237
Hörhammer, I.	966	Jaulent, M.-C.	623, 1227
Hostrup Nielsen, P.	v	Jauregui, I.	1181
Houben, R.	1319	Jeevanandan, N.	766
Howe, L.	1041	Jefferson, E.	1327
Hsu, T.-Y.	1317	Jensen, M.H.	1179, 1413
Hsueh, P.-Y.S.	1006	Jeong, A.	1168
Huang, L.	874	Jeong, I.C.	658, 1066
Hubaux, J.-P.	238, 317, 1161	Jerez, E.	1279
Huber, S.	397	Jesus, R.	663
Hübner, U.	607	Jiang, K.	874
Humbert, M.	238	Jormanainen, V.	833, 833
Hunold, P.	123	Jousselmé, C.	1401
Hurst, S.	1098	Joy, M.	1339
Hüsters, J.	607	Juhra, C.	1297
Husnayain, A.	853	Jung, O.	1291
Hussein, R.	869	Jung, S.	1249
Idiagbonya, E.	1255	Junger, A.	1275
Idiagbonya, J.	1255	Junttila, K.	1177
Ilic, N.	1325	Jury, F.	936
Iliopoulou, D.	509	Kabitzsch, K.	828
Ing Lorenzini, K.	48	Kaiser, L.	1163
Ingenerf, J.	437, 1185	Kakalou, C.	848
Iqbal, U.	1263, 1399	Kaliukhovich, D.A.	287

Kalra, D.	228, 623	Korhonen, O.	1385
Kaltoft, M.K.	532, 562	Koster, Y.	1349
Kamann, C.	597	Kosyrkova, A.	163
Kaminker, D.	818	Kotik, K.	163, 382
Kang, M.	736	Kouame, A.	791
Kao, S.	736	Koutkias, V.	848, 941
Karamanidou, C.	941	Koutsouris, D.	143, 509
Karni, L.	623	Kovalchuk, S.	43, 342
Karopka, T.	463	Kovatchev, P.	552
Karpatkin, H.	658	Kowalski, C.	1245
Katsuki, T.	277, 1289	Kowski, N.	1170
Katzensteiner, M.	272	Kraecher, F.	443
Katzer, C.	1295	Kralevska, K.	1293
Kavanagh, É.	1423	Krämer, A.	1061
Kawohl, C.	1061	Krastev, E.	552
Kern, J.	1185	Kreiner, K.	761
Khemthong, K.	1407	Kreuzthaler, M.	83
Khiewkhern, S.	1343	Kristiansen, M.	879
Khruekarnchana, P.	1357	Kronborg, T.	731
Kikuchi, K.	1347	Krückeberg, J.	1305
Kim Sing, C.	1417	Kuhn, K.A.	68
Kim, D.	736	Kuhns, L.	1365
Kim, H.H.	736	Kujala, S.	966
Kim, J.	736	Kulau, U.	808
Kim, S.M.	1373	Kulenovic, A.	648
Kim, T.R.	1373	Kulikowski, C.A.	1113
Kinnunen, U.-M.	771, 1143	Kuloheimo, A.	1177
Kiourtis, A.	13	Kümmel, M.	158, 597
Kis, J.	1297	Kunimatsu, A.	1187
Kitsuregawa, M.	407	Kuo, A.	1425
Kivekäs, E.	771	Kuo, T.B.J.	1317
Klawunn, R.	1305	Kuroda, T.	718, 1247, 1363
Klein, G.O.	623	Kushniruk, A.W.	1076
Kluge, E.-H.	1094	Kuusisto, A.	1267
Knaup, P.	347, 741	Kwon, H.W.	736
Knowles, S.	936	Kyburz, P.	673
Ko, R.K.L.	1118	Kyriazis, D.	13
Kobrinskii, B.	33	Kyrimi, E.	1239
Koch, S.	643	Kyurkchiev, H.	552
Kockek, P.	1273	Lablans, M.	1185
Kock-Schoppenhauer, A.-K.	1185	Lacoste, S.	1401
Koh, Z.X.	1357	Lafuente-Lafuente, C.	542
Kolditz, T.	28	Lagumdzija-Kulenovic, A.	648
Kolstad, M.	1323	Lakrafl, C.	1313
Konishi, S.	23, 203	Laksono, I.S.	853
Kopanitsa, G.	916, 1379	Laleci Erturkmen, G.B.	623
Kopitowski, K.S.	838	Lamer, A.	218, 247
Kopka, M.	1209	Lampoukas, E.	143
Kopka, R.	1209	Lamy, J.-B.	1313, 1329

Lamy, J.B.	791	Lindsay, H.	1417
Larbi, D.	796	Lins, C.	653
Lardani, L.	1201	Lipprandt, M.	1419
Largiadèr, C.R.	618	Liu, J.	1066
Larouche, A.	1423	Liu, N.	1357
Larsen, T.K.	1179	Liyanage, H.	1339, 1369
Laurent, G.	218	Ljosland Bakke, S.	443
Laursen, S.H.	1403	Llanos, L.C.	432
Lavaire, B.	108	Lo, M.	791
Lawrence, A.K.	1156	Löbe, M.	392
Le Magny, D.	108	Loftsgard, K.	1417
Le Sueur, H.	387	Lohr, C.	28
Leblanc, A.	1423	Looten, V.	1133
Lebre, R.	663	Lopez Osornio, A.	1337
Lee, C.W.-C.	1417	López Osornio, A.	818, 1011, 1377
Lee, E.	1249	Lopez, D.M.	352
Lee, J.	1249	Lopez, E.	1377
Lee, Y.J.	1253	López, E.	818, 1011
Leeming, G.	1103	Lopez, V.	173, 879
Légaré, F.	1423	Lopez-Campos, G.H.	1195, 1331
Legrand, B.	247	López-Úbeda, P.	292
Lehne, M.	8	Lorent, M.	458
Leib, S.L.	1163	Losilla, G.	1170
Leibowitz, K.	1041	Lovis, C. v, vii, 48, 198, 208, 322, 894,	
Leichtle, A.	1161	1016, 1081, 1098, 1361	
Leis, A.	921	Lu, T.-C.	1357
Lemanska, A.	1369	Ludmann, D.	1395
Lemordant, P.	547	Ludvigsen, A.-E.	453
Lenne, C.	108	Lueg, C.	613
Lépinay, F.	1423	Luna, D.	188, 668, 776, 901, 1181,
Les, T.	458	1217, 1271, 1279, 1375	
Leslie, H.	443	Lungeanu, D.	1303
Lesselroth, B.	1076	Lüpkes, C.	233, 653
Levi, D.	582	Lutterotti, A.	168
Lewi, M.	367	Luzi, D.	223, 484
Lewis, S.	1367	Ma, H.-M.	1357
Li, Y.-C.J.	153, 1241	Mac Culloch, P.	818
Li, Z.	252	MacCulloch, P.	1377
Liaskos, J.	1197	Machin, M.	1367
Lichtner, V.	1405	Madan, S.	1371
Liesche-Starnecker, F.	1207	Madden, R.	1409
Lilja, M.	623	Madiot, P.-E.	108
Lillo-Lelouet, A.	1227	Maduri, C.	1006
Lim Choi Keung, S.N.	623	Maeder, A.	577, 678
Lim, M.	1417	Maffioletti, S.	1170
Lin, M.-C.	1317, 1355	Magdalinou, A.	1299
Lindgren, M.	833	Magoev, K.	1379
Lindman, P.	623	Mahmut, E.E.	357
Lindoerfer, D.	262	Maid, J.J.	818, 1011

Maignant, G.	1353	Mayer, M.A.	786
Majd, M.	73	Mayer, M.-A.	921
Majeed, R.W.	138	Mazhari, R.	1305
Makaram, N.	1219	McBeth, J.	946
Makashova, E.	163	McDuffie, M.	18
Makhlysheva, A.	148	McLachlan, S.	1239
Makino, M.	277, 1289	McNicol, I.	567
Mamandipoor, B.	73	McQuaid, K.	1195
Mamuye, A.	1251	Mea, V.d.	1409
Manabe, S.	23, 203	Meder, B.	1061
Mandirola Brieux, H.	1377	Medlock, S.	981
Mandirola Brieux, H.F.	1337	Mehrabi, M.	713
Mandloi, S.	1205	Meier, L.	613
Mangalji, A.	1417	Meineke, F.A.	392
Mansmann, U.	262	Meinshausen, N.	1163
Manta, O.	509	Meredith, J.	567
Mantas, J.	1197, 1299	Metzker, O.	1379
Manyakov, N.V.	287	Meyer, R.	312
Manzotti, M.	1211	Mezinska, S.	1138
Maramis, C.	941	Michalopoulos, G.	53
Marangone, A.	587	Michel, P.-A.	884
Maranhao, P.	1315	Michel-Backofen, A.	138, 1295
Marceglia, S.	1201	Michelsen, T.	653
Marchio, V.	1201	Michielin, O.	1161
Marcilly, R.	218	Micucci, D.	693
Marguerie, C.	623	Mielke, C.	746
Marinello, R.	522	Miettinen, S.	572
Marino, F.	1161	Mihalas, G.I.	1113
Markiewicz, M.B.	1011	Mihara, N.	1036
Markiewicz, T.	458	Milic, N.	1325
Markota, A.	1273	Milicevic, O.	1325
Maršálová, K.	1221	Milin-Lazovic, J.	1325
Marsch, S.	1163	Millius, S.	58
Marschollek, M.	479, 808	Mingarelli, V.	484
Marsh, D.W.R.	377	Minoletti, S.	901, 1181
Martignene, N.	247	Miotti, S.	484
Martin, D.K.	1215	Miribel, A.	198
Martínez, B.	776	Misbach, M.	317, 1161
Martinez, S.	1259	Mitsutake, N.	407
Martinuzzi, A.	1409	Modersohn, L.	28
Mascia, C.	443	Moen, A.	926
MASTERplans Consortium	387	Mohd Yusof, M.	906, 1036
Masud, J.H.B.	1355	Moldes, R.	1377
Matos, S.	93	Molina, C.	402
Matsopoulos, G.	143	Molina, J.	1170
Matsumura, Y.	23, 203, 1036	Moll, J.	1021, 1056
Matthies, F.	392	Moncada-Torres, A.	307
Maurer, J.	1151	Monkman, H.	1026, 1076
Mavrogiorgou, A.	13	Montandon, L.	1189

Monteiro, E.	1183	Niyitegeka, D.	412
Mora, S.	1371	Niyu, H.	1253
Moreau-Gaudry, A.	108, 1215	Nkhom, D.	1263
Moreno, D.	1170	Nøhr, C.	766, 956, 1026
Morgan, G.	678	Nunes, P.	663
Mossuz, P.	108	Nurhussien, F.	1251
Mottaz, A.	884	Nyssen, M.	803
Mottin, L.	884	Nyulas, C.I.	1409
Motulsky, A.	1001, 1423	O'Mahony, G.	1195
Moz, M.	73	O'Sullivan, D.	327
Mucha, T.	113	Oberfeld, J.	1297
Muehlbauer, V.	1207	Oberije, C.	1319
Müller, H.	58, 302	Odou, P.	683
Mulligan, N.	173	Oehm, J.	592
Mullin, S.	823, 1381	Oemig, F.	1089
Muñoz, A.	1243	Oertl, M.	1207
Muro, N.	517	Oesterle, S.	1170
Musen, M.A.	1409	Oehlmeier, C.	1297
Mylonopoulou, V.	1385	Ohtera, S.	1247, 1363
N'Dja, A.	367	Okamoto, K.	1247, 1363
Nakao, A.	638	Okhmatovskaia, A.	858
Nakashima, N.	1347	Okoniewski, M.	1170
Nakayama, M.	1321	Olakotan, O.	906
Nambu, M.	1363	Oleynik, M.	83, 292
Namdar, M.	198	Olive, F.	108
Namli, T.	1291	Oliveira, J.L.	1183, 1391
Nanni, O.	367	Omnes, S.	453
Nap, H.H.	1081	Ong, M.	1357
Napel, H.t.	1409	Ono, M.	277, 1289
Naroo, G.Y.	1357	Orcaizaguirre, N.	1287
Nassirian, A.	158	Oreve, M.-J.	1401
Navarro, G.	367	Orloff, M.	183
Ndiaye, M.-A.	1423	Orsi, C.	297
Ndiaye, N.	1387	Óskarsdóttir, M.	996
Neeser, A.	1170	Osmani, V.	73
Ness, S.	287	Otero, C.	1181
Neuraz, A.	432	Otten, H.	1395
Newby, L.K.	337	Otter, S.	1369
Nghiem, D.	542	Pagani, J.-L.	1163
Nguyen, P.-A.	1241	Pagonis, D.	108
Nguyen, V.	474	Palanisamy, R.	1309
Nicola, S.	756	Palojoki, S.	771
Nicolás, J.M.	1215	Pan, J.Z.	1327
Nikou, G.	1351	Pandina, G.	287
Nind, T.	1327	Panizza, E.	1265
Nishihara, Y.	1323	Pantazoglou, E.	1395
Nishikitani, M.	1347	Papassotiropoulos, A.	1168
Nissinen, S.	1281	Pape-Haugaard, L.B.	v, vii
Nitulescu, A.	128	Paralič, J.	537

Park, S.	1249	Pietruski, P.	1209
Park, Y.R.	736, 1249	Pilutti, D.	1223
Parrend, P.	1353	Pireddu, L.	443
Parvanova, I.	1066, 1235	Pires, F.	103
Pasche, E.	884	Pischek-Koch, K.	607
Paskal, A.M.	1209	Pitoglou, S.	143
Paskal, W.	1209	Plagianakos, V.	1307
Pathirannahelage, S.	1339	Planell-Morell, P.	469
Patte, M.	1197	Plazzotta, F.	776, 931
Paturel, I.	108	Ploessnig, M.	1291
Pauli-Magnus, C.	1151	Ploner, N.	991
Paulus, W.	1170	Poly, T.N.	153, 1241
Pazos, A.	1183	Pözl, G.	761
Pearson, C.	1365	Pomares-Quimbaya, A.	292
Pecoraro, F.	223, 484	Popescu, M.H.	297
Pedernera, F.A.	818, 1011	Potapov, A.	163, 382
Pedersen, M.S.	951	Pou, A.	1265
Pedersen, S.B.	951	Pourabbas, E.	223
Pedrera, M.	78, 1243	Pradervand, S.	1161
Pedretti, A.	776	Prados-Suarez, B.	402
Pedrosa, A.	103	Prasser, F.	8, 68, 193
Peek, N.	387, 1367	Prebet, C.	628
Pek, P.P.	1357	Preisig, M.	1168
Pelayo, S.	218	Prenassi, M.	1201
Pella, D.	537	Prgomet, M.	1405
Peltonen, L.-M.	1177	Prior, F.	183
Peltzer, K.	1407	Priou, S.	98
Peña-Yañez, C.	402	Probst-Hensch, N.	1168
Peng, M.	88	Prokop, G.	1207
Penner, I.-K.	168	Prokosch, H.-U.	991, 1061
Penning, M.	118	Pronkina, T.	163, 382
Penning, M.L.	183, 337	Provaznik, I.	1221
Peper, F.E.	1341	Prowle, J.	377
Pereira, A.M.	1315	Pryzsucha, M.	607
Pereira, H.	367	Pugin, J.	1163
Pérez-Rey, D.	78	Pujadas, G.	1011
Perisse, D.	1401	Puscoci, S.	723
Petersen, C.	1123	Pusztová, L.	537
Petersen, L.S.	956	Puts, S.	1319
Petitgand, C.	1001	Qazi, H.	53
Petre, M.	723	Quan, H.	88
Petrossians, P.	552	Quantin, C.	213
Peute, L.	1071, 1257, 1277	Quentin, P.	63
Pfeifer, Î.	761	Quesnel-Barbet, A.	1353
Pfuhl, G.	894	Quintana, G.N.	1283
Philippe, A.	1401	Raahauge, A.	848
Philips, M.	618	Raben, D.	848
Picco, G.	879	Radix, A.	1365
Pieper, C.	337	Rahimi, S.A.	1423

Raine, R.	469	Röhrig, R.	1128, 1419
Raineri, M.	367	Rojas Mezarina, L.	474
Raisaro, J.L.	238, 317, 1161	Rolandi, F.	818, 1011
Rajalahti, E.	1143	Romanchikova, M.	1369
Rajput, V.K.	532	Romand, M.	322
Rance, B.	432	Ronzano, F.	921
Randine, P.	693	Rosat, N.	1161
Ranegger, R.	38	Rosati, S.	1193
Rapin, M.	751	Rosset, S.	432
Rapisarda, R.	901	Rossi Mori, A.	223
Rätsch, G.	1163	Roumen, C.	1319
Ratti, M.F.G.	776	Rubinstein, A.	818, 1011
Ravey, J.-N.	108	Ruch, P.	312, 884
Recalde, I.	582	Ruiz Garretas, V.	1098
Redeker, G.	688	Ruiz, A.A.	352
Rees, M.	936	Ruotsalainen, P.	1089
Régis, C.	1001	Sa, J.	1161
Reinecke, I.	158, 1393	Saarni, L.	1143
Reinhardt, I.	262	Saccilotto, R.	1151
Reiso, H.	453	Sakilay, S.	1381
Relander, T.	833	Salamanca, A.	1243
Rexhepi, H.	1056	Salanterä, S.	1177
Reymen, B.	1319	Salehi, A.	713
Rho, V.	879	Salehi, E.	713
Ricci, A.	1081	Salgado, C.	1243
Ricci, F.L.	223, 484	Samer, C.	48, 1098
Richard, A.	894	Sampri, A.	387
Richter, J.	607	Samudrala, R.	1205, 1331
Ridao, M.	1397	Sandulescu, V.	723
Rinn, B.	1170	Santavirta, J.	1267
Rissanen, S.	1108	Sanz, C.	1243
Rivera-Romero, O.	911	Sanz, F.	921
Rizk, B.	58	Sarafidis, M.	143
Rizzato Ledes, D.A.	818, 1011, 1265, 1337, 1377	Saranto, K.	771, 1108, 1267
Robel, L.	1401	Sarigul, B.	623
Roberge, J.	1423	Sariyar, M.	397, 494, 618, 673
Robert, A.	198, 208	Sasseville, M.	1423
Roberts, K.	813	Sato, J.	407
Robertson, D.	1327	Sato, K.	894
Robins, D.	1066	Sato, Y.	1347
Rocca, M.	961	Sauleau, E.-A.	1353
Rochat, J.	322, 894, 1081, 1098, 1361	Savic, M.	1325
Rochoy, M.	683	Sayers, K.	1170
Rodier, S.	843	Sbodio, M.	173, 879
Rodríguez Tablado, M.	1333	Scandurra, I.	1031
Rodriguez, G.	1291	Schachner, B.	188
Rodriguez, M.	587	Schaer, R.	302
Rodriguez-Blanco, O.	786	Schamne, F.K.	1233
		Schepers, J.	8

Schermer, R.	981	Silva Valencia, J.	474
Scheuermann, B.	8	Silva, J.F.	93
Schiavoni, V.	332	Silva, L.B.	1183
Schmeer, R.	1305	Silva, P.	946
Schmidt, T.	1026, 1229	Simon, M.	1133
Schmitt, K.-U.	673	Simón, M.	931, 1375
Schnall, R.	1365	Simonsen, J.	971
Schneider, B.	463	Smaradottir, B.F.	453
Schoppmann, P.	8	Smith, M.	1279
Schreier, G.	761	Smith, N.A.S.	1369
Schrenzel, J.	1163	Sohn, H.-S.	736
Schröder, D.	1061	Soika, K.	123
Schulz, S.	83, 292, 352	Soko, C.J.	1263
Schumacher, S.	443	Sommer, J.	776, 931, 1375
Schummer, G.	108	Song, K.J.	1357
Schüpbach, R.	1163	Song, L.	1255
Schut, M.	257	Soula, J.	1353
Schwarz, D.	1221	Spechbach, H.	527, 1421
Schwede, T.	1170	Speranza, C.D.	818, 1011, 1265, 1377
Schweig, M.	1061	Speranza, M.	1401
Sciarreta, S.	1285	Spetsarias, A.	143
Sconfienza, L.	242	Spier, R.	1076
Scott, P.	v	Springer, A.	1203
Secher, P.H.	1359	Staccini, P.	602, 1353
Sedki, K.	63, 1313	Stamatopoulos, K.	941
Sedlmayr, M.	158, 597, 1393	Stanisavljevic, D.	1325
Segarra, C.	332	Stäubert, S.	392, 463
Segrave-Daly, J.	173, 879	Stausholm, M.N.	1359
Sellemann, B.	607	Stein, J.	658
Selter, L.	1156	Stein, R.	1377
Semeraro, D.	1391	Steinke, D.	1295
Sengstag, T.	1170	Steinmann, O.	986
Sepehri, A.A.	178	Stekhoven, D.J.	884
Seppälä, T.	113	Stieltjes, B.	1168
Séroussi, B.	542, 628, 698	Štiglic, G.	1273
Serrano, P.	78, 1243	Stockinger, H.	1170
Seth-Smith, H.	397	Stockton-Powdrell, C.	1367
Shaban-Nejad, A.	448	Stöhr, M.R.	138
Sharifullin, F.	33	Stoicu Tivadar, V.	357
Sherlock, J.	1369	Stoicu-Tivadar, L.	vii, 128, 756
Sherman, M.	623	Stojanovic, M.	843
Shibata, T.	1187	Storck, M.	592
Shifrin, M.	163, 382	Storf, H.	986
Shimizu, Y.	1187	Störk, S.	1415
Sick, R.	1297	Strohmeier, F.	1291
Sicotte, C.	1423	Stroth, A.M.	123
Siegemund, M.	1163	Strunina, U.	382
Sigg, S.	808	Strunina, Y.	163
Sijm-Eeken, M.	1071	Stucke, O.	1311

Studer, T.	168	Thun, S.	8
Su, E.C.-Y.	853	Tiainen, M.	1143
Su, K.	1255	Tilahun, B.	1251
Sudlow, C.	1327	Tippenhauer, K.	618
Sugimoto, K.	203	Tocimáková, Z.	537
Sugiyama, O.	718, 1247, 1363	Todorovic, M.	367
Sulis, E.	522	Tokumasu, R.	1289
Sultana, N.	1347	Tomczak, A.	443
Sundgren, M.	367	Torrejón, A.	786
Sung, M.	1249	Torrens, M.	921
Suominen, T.	1267	Torres, J.	517
Surkont, J.	1170	Tortajada, S.	864
Suwanlee, S.R.	1407	Traore, L.	623
Suzuki, A.	277, 1289	Trifan, A.	1391
Swaminathan, R.	1219, 1225, 1269, 1309	Troncoso, C.	238
Swoboda, W.	1207	Troncoso-Pastoriza, J.R.	238, 317, 1161
Syed, M.	183, 337	Troqe, R.	527
Syed, S.	183, 337	Tsopra, R.	63, 133, 1313, 1329
Syeda, H.B.	183	Tsourakis, N.	527
Sykes, C.	1409	Tsukanova, T.	163
Sylvestre, E.	18, 547	Tsuru, S.	638
Takahashi, D.	23	Tu, S.W.	1409
Takeda, T.	23, 203, 1036	Tubert-Bitter, P.	213
Tamamoto, T.	638	Tudorache, T.	1409
Tamburis, O.	484	Tulo, S.K.	1225
Tanaka, H.	1357	Turnbull, N.	1407
Tanila, T.	708	Turner, A.M.	1213
Tanizaki, K.	638	Turner, J.	327
Tapuria, A.	228	Turner, P.	1123
Tarousi, M.	143	Tytherleigh, C.	98
Tasneem, R.	1347	Udalov, Y.D.	504
Tcharaktchiev, D.	552	Udsen, F.W.	1403
Tchuenté, J.	1423	Ugon, A.	63, 133
Teckert, Ö.	633	Ulgu, M.M.	557
Temkin, A.	372	Ulrich, H.	1185
Ten Teije, A.	307	Umemoto, K.	407
Tenhunen, H.	113, 572, 708	Uribe, G.A.	352, 1089
Tensen, E.	1257, 1277	Utunen, H.	1311, 1387
Terrasa, S.	1383	Uva, P.	443
Terrasa, S.A.	838	Valdivieso, B.	864
Terriza, A.	1243	Valeri, F.	1193
Testoni, S.	367	Valette, M.	63
Thevenon, J.	108	Valinoti, M.	776
Thew, S.	1103	van Buggenum, J.	1257
Thiemann, V.S.	1128	van der Heijden, B.	1319
Thomas, S.A.	1369	van der Heijden, J.	1277
Thongkum, W.	1343	Van Der Veer, S.N.	936, 946
Thouly, C.	58	van der Velde, N.	257

van Gool, C.	1409	Wilk, M.	377
Van Laere, S.	803	Willers, J.	1151
Van Maaren, M.	307	Williams, R.	173
van Sinderen, F.	1277	Williams, T.	1199
van Velsen, L.	1046	Winter, A.	vii, 392, 463
Varela, C.	1243	Witkamp, L.	1257, 1277
Varghese, J.	889	Witteman, H.O.	1423
Värri, A.	1143	Wlodarski, P.K.	1209
Vasilyeva, I.V.	504	Woldaregay, A.Z.	894
Velupillai, S.	98	Wolf, M.	808
Verdoy, D.	623	Wollschlaeger, B.	828
Vernugopan, V.	98	Wolters, T.	233
Vestergaard, P.	1403, 1413	Wong, A.	53
Veugen, J.	1319	Wood, S.	1076
Viani, N.	98	Wu, H.	1327
Viernes, B.	474	Wübbenhorst, O.	233
Vilar, R.	864	Wüst, T.	1170
Villena, F.	347	Wüthrich, D.	397
Visentin, M.	1201	Yahagi, N.	638
Vogel, S.	607	Yakovlev, A.	1379
Vogelsang, L.	8	Yamaguchi, J.	23
Volij, C.	1261	Yamaguchi, N.	1323
Vollenweider, P.	1168	Yamahata, A.	203
Volpi, M.	1383	Yamamoto, G.	718, 1247, 1363
Von Tottleben, M.	623	Yamamoto, T.	1247
Vuadens-Lehmann, A.	1421	Yamamoto, Y.	23
Wache, S.	607	Yang, H.-C.	1241
Wacker, J.	751	Yang, H.C.	153
Wada, S.	203	Yanishevskiy, S.	1379
Wagner, T.O.F.	986	Yigzaw, K.Y.	148
Walden, A.	337	Yokota, F.	1347
Walther, D.	986	Yoo, S.	1373
Wang, M.	1327	Yuan, C.	252
Wang, P.	337	Yuksel, M.	623
Wang, Y.	843	Zamboni, N.	1163
Wang, Y.-H.	1241	Zana, I.	1227
Wang, Z.	337	Zanetti, G.	443
Warnecke, T.	889	Zarubina, T.V.	504
Warren, J.	976	Zbinden, R.	1163
Wartenburger, R.	1170	Zebbities, S.	607
Watkins, M.	474	Zeng, Q.	1327
Weber, M.	3	Zeppernick, T.	808
Weber, P.	v	Zhao, U.	1387
Weisburd, B.	372	Zheng, J.	1071
Weng, C.	252	Zinkernagel, A.	1163
Were, M.C.	1301	Zozus, M.	183, 337, 961, 1203
Westbrook, J.I.	1405	Zozus, M.N.	1199
Whitehead, N.	567	Zubke, M.	272
Whitlock, T.	936	Zylberberg, P.	1401